

2021

Gepi tanulas jegyzokonyv

JEGYZOKONYV

BOZSIK ARMAND VIKTOR

Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev

A valasztott UCI adatbazisok (3 db)

- 1. Breast Cancer Wisconsin (Diagnostic) Data Set
- 2. Breast Cancer Coimbra Data Set
- 3. Abalone

Mellekelve: UCI/x/db mappa.

Az elkeszített saját adatbazisok (2 db)

- 1. Bonyolultabb, spiral-szerű
- 2. Egyszerűbb

Megjegyzés: Az egész projektet a jegyzőkönyvvel egyező nevet viselő ZIP-ben adom be. A ZIP tartalmazza jelen jegyzőkönyv pontos másolatát is!

A projekt itt található: (Amennyiben CooSpace-re nem fer fel)

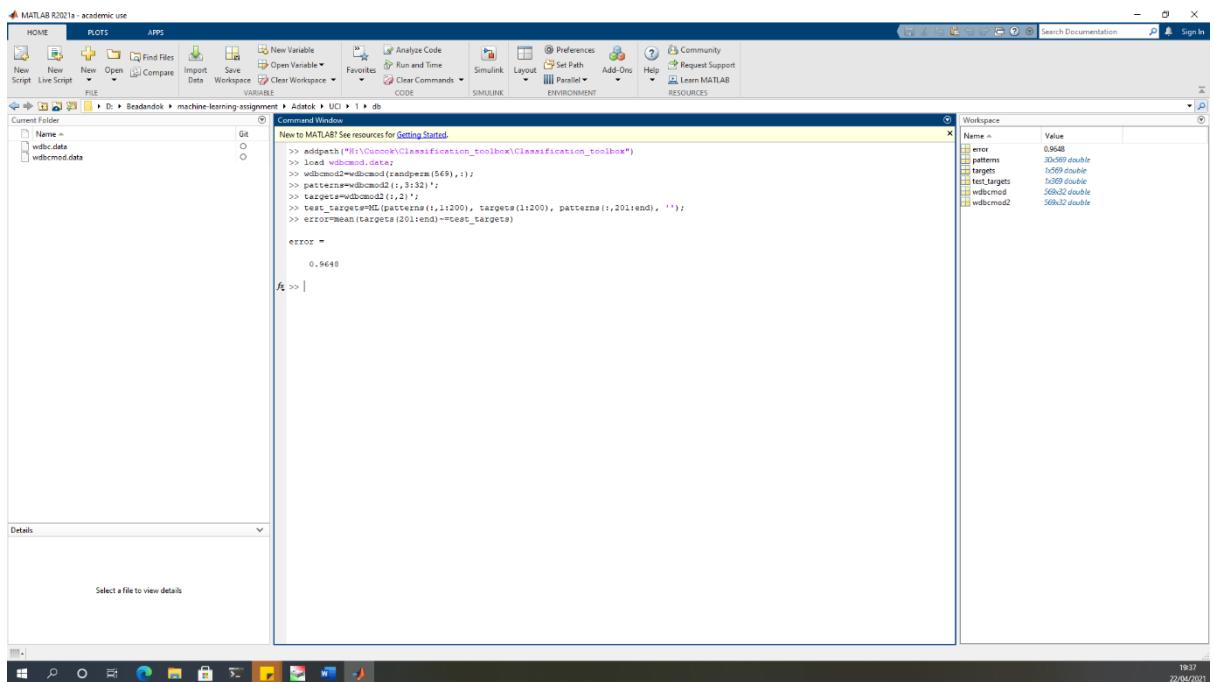
Keszítette: Bozsik Armand Viktor
 Programtervező Informatikus MSc (levelező) szakos hallgató
 2020/21 II. felev
UCI adatbazisok

1. UCI adatbazis

Adatok	
Neve:	Breast Cancer Wisconsin (Diagnostic) Data Set
Sorainak szama:	569
Jellemzőinek szama:	32
Osztalyainak szama:	2
Elvezett alakitasok (ha szukseg es volt):	Lementettem a kozzettet data kiterjeszes fajlt (wdbc.data), majd atalikitottam (wdbc-mod.data). A cimkeosztalyok M, B ertekeit atalikitottam 1-2-re, replace all-lal). Betoltottem a MATLAB-ba a tanult load parancsal.
Sikertelen futas oka, leirasa (ha volt ilyen):	Lent (ha volt ilyen)

A futasrol készült kepernyokepek (Rajtuk a hibaszazalek)

1. ML:



Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev

2. ML_diag:

The screenshot shows the MATLAB R2020a interface. The Command Window displays the following code and its execution results:

```
>> addpath('C:\Users\user\OneDrive\Classification_toolbox\Classification_toolbox')
>> load wbdcmod.data;
>> wbdcmod2=wbdcmod1(randperm(569),:);
>> patterns=wbdcmod2(:,1:32)';
>> targets=wbdcmod2(:,2)';
>> test_targets=ML_diag(patterns(:,1:200), targets(1:200), patterns(:,201:end), '');
>> error=mean(targets(201:end)-test_targets)

error =
0.9648

f2 >> |
```

The Workspace browser on the right side shows the following variables:

Name	Type	Value
error	double	0.9648
patterns	double	30x32 double
targets	double	1x329 double
test_targets	double	30x32 double
wbdcmod	double	30x32 double
wbdcmod2	double	30x32 double

3. Nearest_neighbor:

The screenshot shows the MATLAB R2020a interface. The Command Window displays the same code as the previous screenshot, but with a different error value:

```
>> addpath('C:\Users\user\OneDrive\Classification_toolbox\Classification_toolbox')
>> load wbdcmod.data;
>> wbdcmod2=wbdcmod1(randperm(569),:);
>> patterns=wbdcmod2(:,1:32)';
>> targets=wbdcmod2(:,2)';
>> test_targets=Nearest_Neighbor(patterns(:,1:200), targets(1:200), patterns(:,201:end), 7);
>> error=mean(targets(201:end)-test_targets)

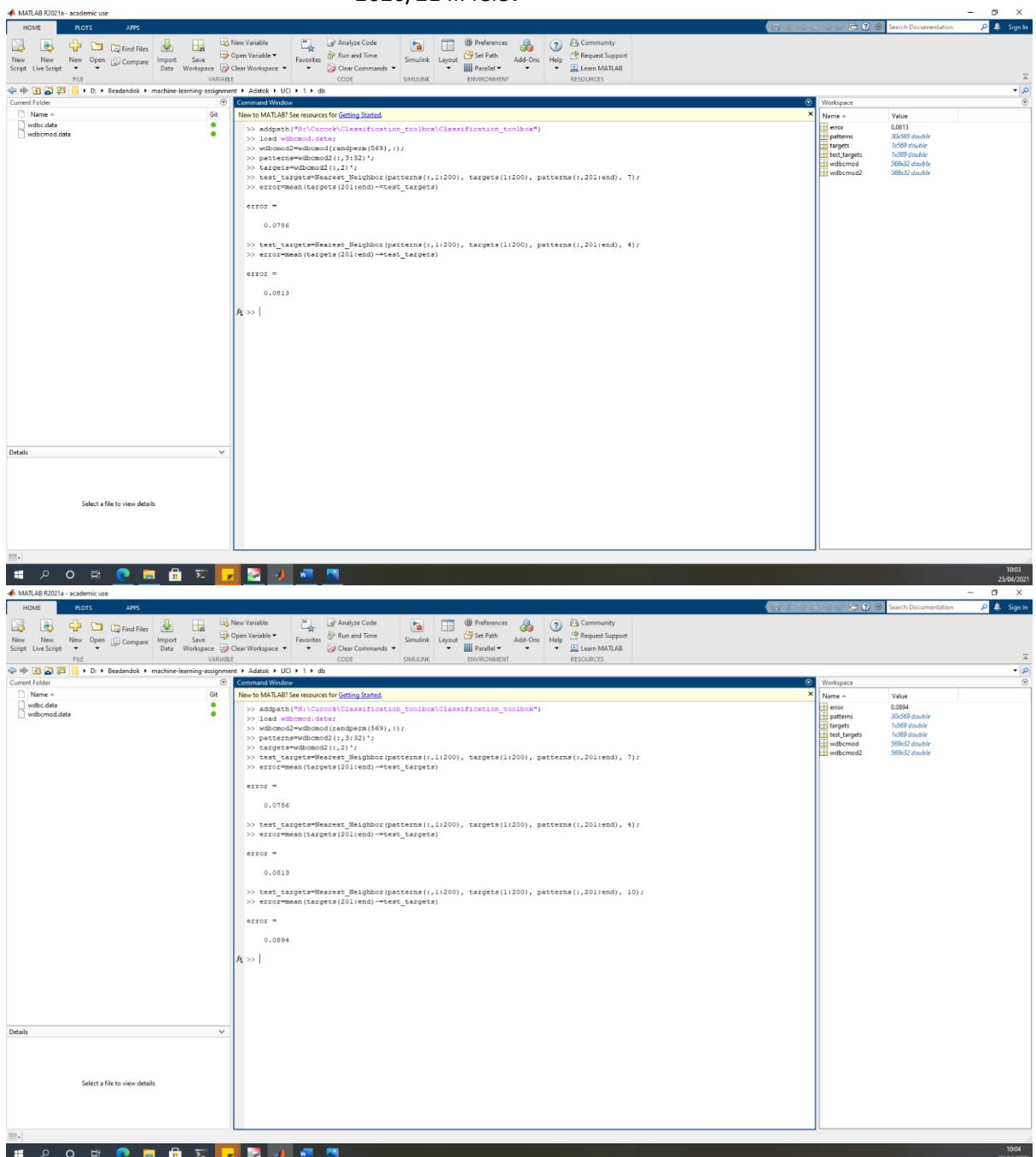
error =
0.0786

f2 >> |
```

The Workspace browser on the right side shows the following variables:

Name	Type	Value
error	double	0.0786
patterns	double	30x32 double
targets	double	1x329 double
test_targets	double	30x32 double
wbdcmod	double	30x32 double
wbdcmod2	double	30x32 double

Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev



```

>> addpath('H:\Cucco\Classification_toolbox\Classification_toolbox')
>> load wbdcmod.data;
>> wbdcmod2='wbdcmod1(randperm(569),:)';
>> patterns=wbdcmod2(:,1:32);
>> targets=wbdcmod2(:,33:32)';
>> test_targets=wbdcmod2(:,1:2);
>> test_error=mean(targets(201:end)==test_targets);
>> error=mean(targets(1:200), targets(1:200), patterns(:,201:end), 7);

error =
0.0786

>> test_targets=Nearest_Neighbor(patterns(:,1:200), targets(1:200), patterns(:,201:end), 4);
>> error1=mean(targets(201:end)==test_targets);

error =
0.0813

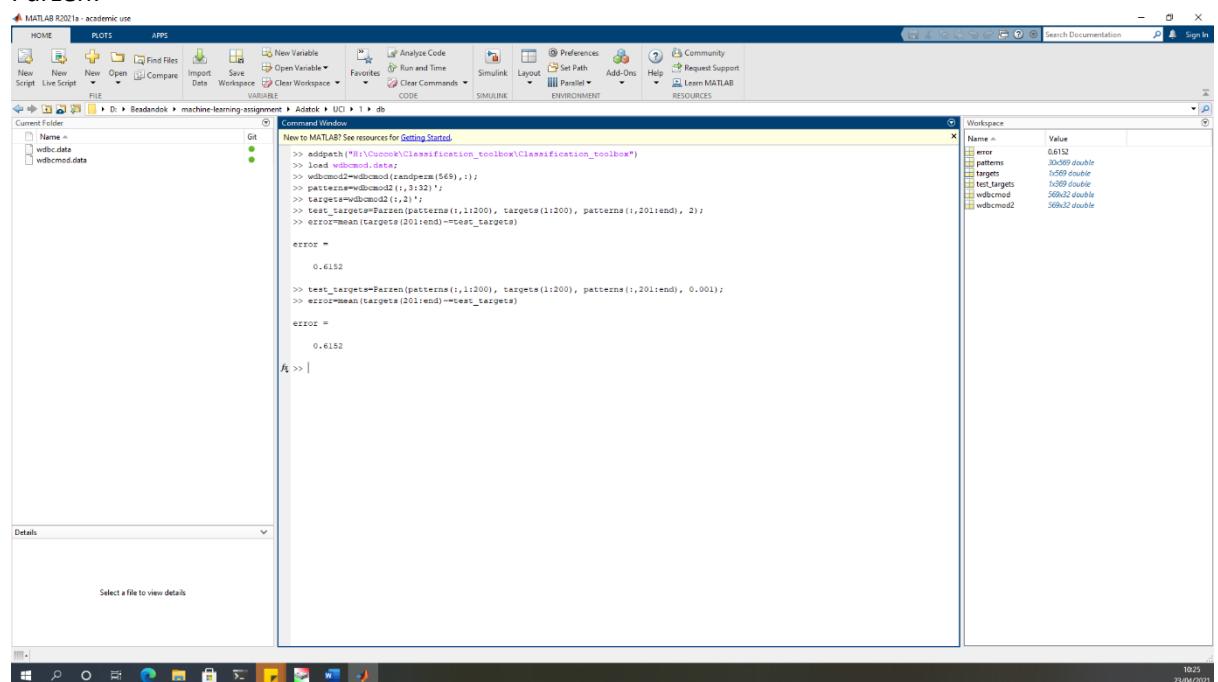
>> test_targets=Nearest_Neighbor(patterns(:,1:200), targets(1:200), patterns(:,201:end), 10);
>> error2=mean(targets(201:end)==test_targets);

error =
0.0894

```

Keszítette: Bozsik Armand Viktor
 Programtervező Informatikus MSc (levelező) szakos hallgató
 2020/21 II. felev

4. Parzen:



The screenshot shows the MATLAB R2021a interface. The Command Window displays the following code and its execution results:

```

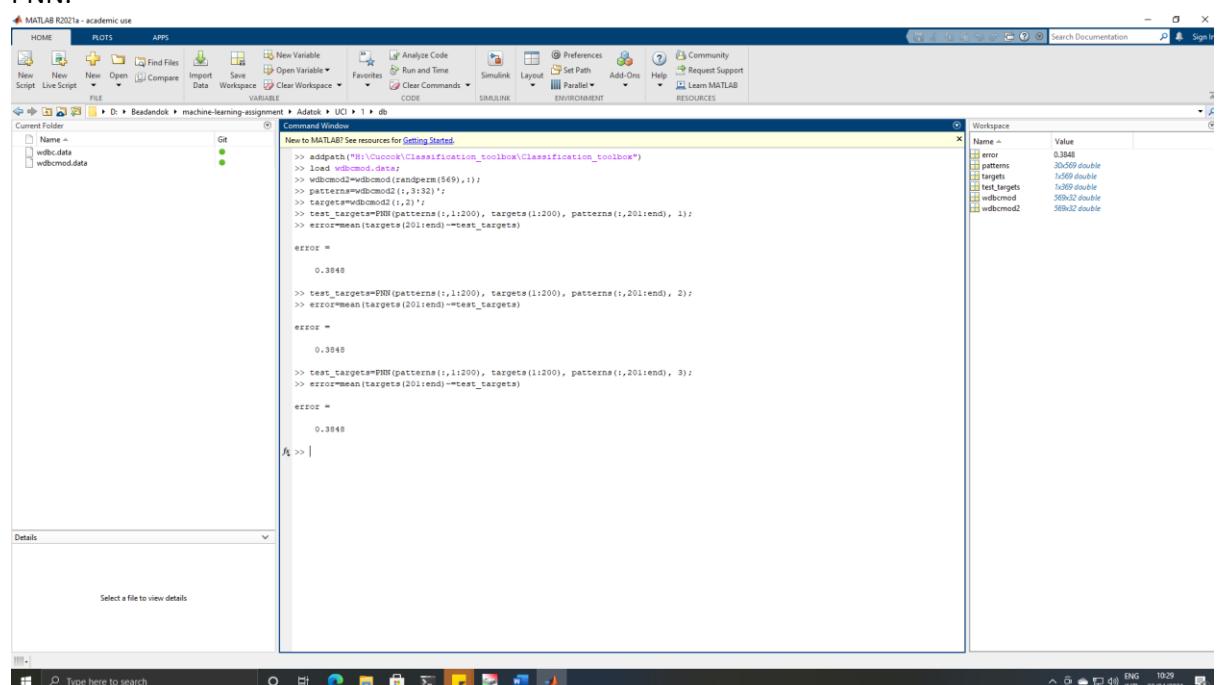
>> addpath('H:\U\codok\Classification_toolbox\Classification_toolbox')
>> load wbcmd1.data;
>> patterns=wbcmd1(:,1:32);
>> targets=wbcmd1(:,33);
>> test_targets=Parzen(patterns(1:1200), targets(1:1200), patterns(1:201:end), 2);
>> error=mean(targets(201:end))-test_targets
error =
0.6152
>> test_targets=Parzen(patterns(1:1200), targets(1:1200), patterns(1:201:end), 0.001);
>> error=mean(targets(201:end))-test_targets
error =
0.6152
f1 >> |

```

The Workspace browser on the right side shows the following variables:

Name	Type
error	0.6152 double
patterns	1x569 double
targets	1x569 double
test_targets	1x369 double
wbcmd1	369x32 double
wbcmd2	369x32 double

5. PNN:



The screenshot shows the MATLAB R2021a interface. The Command Window displays the following code and its execution results:

```

>> addpath('H:\U\codok\Classification_toolbox\Classification_toolbox')
>> load wbcmd1.data;
>> patterns=wbcmd1(:,1:32);
>> targets=wbcmd1(:,33);
>> test_targets=PNN(patterns(1:1200), targets(1:1200), patterns(1:201:end), 1);
>> error=mean(targets(201:end))-test_targets
error =
0.3948
>> test_targets=PNN(patterns(1:1200), targets(1:1200), patterns(1:201:end), 2);
>> error=mean(targets(201:end))-test_targets
error =
0.3948
>> test_targets=PNN(patterns(1:1200), targets(1:1200), patterns(1:201:end), 3);
>> error=mean(targets(201:end))-test_targets
error =
0.3948
f1 >> |

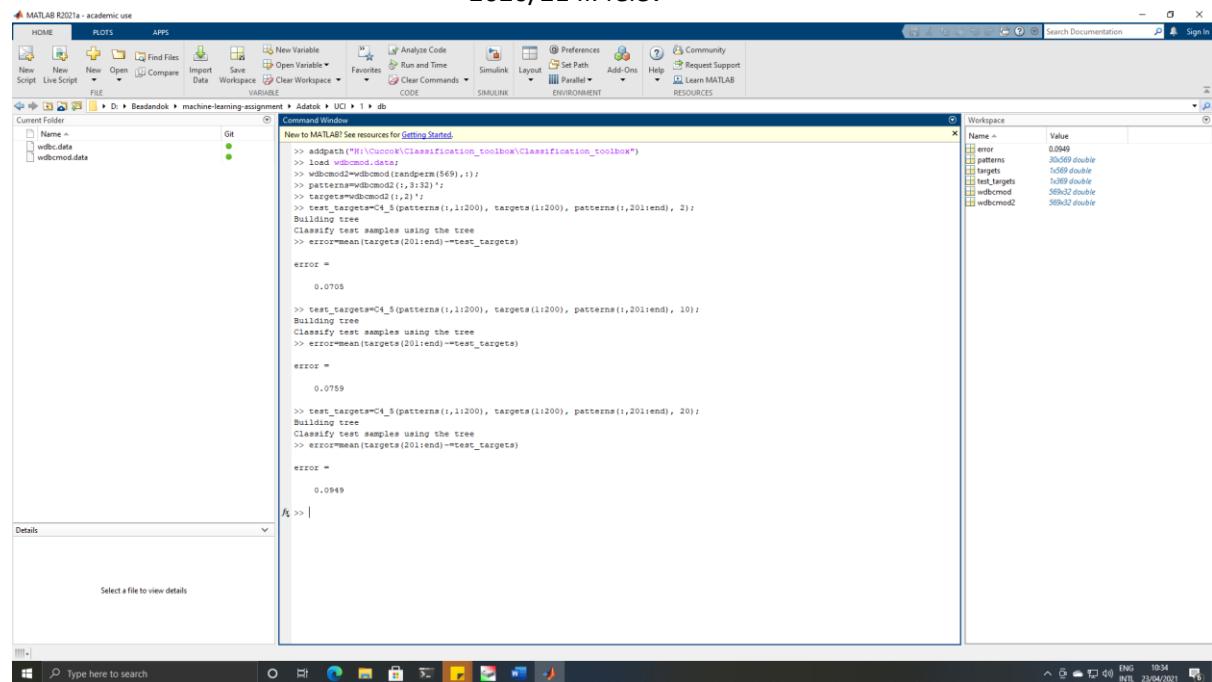
```

The Workspace browser on the right side shows the following variables:

Name	Type
error	0.3948 double
patterns	369x32 double
targets	1x369 double
test_targets	1x369 double
wbcmd1	569x32 double
wbcmd2	369x32 double

6. C4_5:

Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev

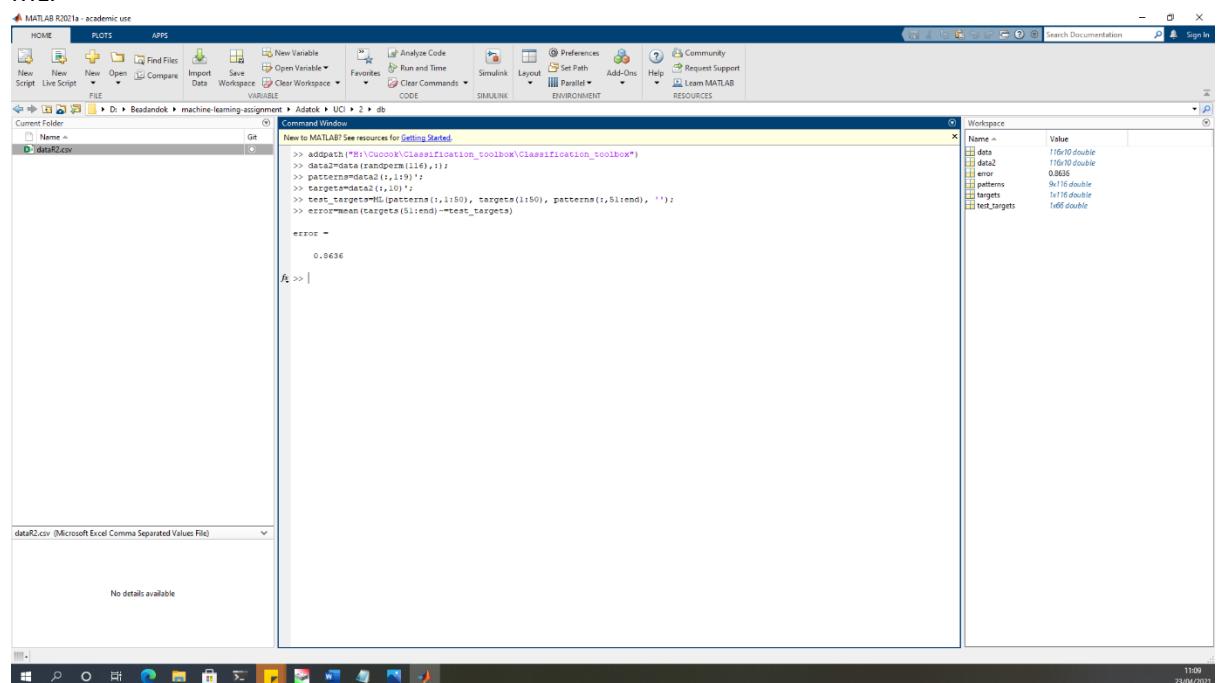


2. UCI adatbazis

Adatok	
Neve:	Breast Cancer Coimbra Data Set
Sorainak szama:	116
Jellemzőinek szama:	10
Osztalyainak szama:	2
Elvegzett atalakitasok (ha szukseges volt):	Lementettem a kozzettett CSV kiterjesztese fajlt (dataR2.csv), majd betoltottem a MATLAB-ba a Help alapjan, grafikusan (table helyett numeric matrixot valasztva): Import Spreadsheets - MATLAB & Simulink - MathWorks United Kingdom
Sikertelen futas oka, leirasa (ha volt ilyen):	Lent (ha volt ilyen)

A futasrol készült kepernyökepek (Rajtuk a hibaszazalek)

1. ML:



Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev

2. ML_diag:

The screenshot shows the MATLAB R2020a interface. The Command Window displays the following code and its execution results:

```
>> test_targets=ML_diag(patterns(:,1:50), targets(1:50), patterns(:,51:end), 1);
>> error=mean(targets(51:end))-test_targets;
error =
0.9242
f1 >>
```

The Workspace browser on the right side shows the following variables:

Name	Type	Value
data	1x10 double	
data2	1x10 double	
error	0.9242	
patterns	9x16 double	
targets	1x16 double	
test_targets	1x10 double	

3. Nearest_neighbor:

The screenshot shows the MATLAB R2020a interface. The Command Window displays the following code and its execution results:

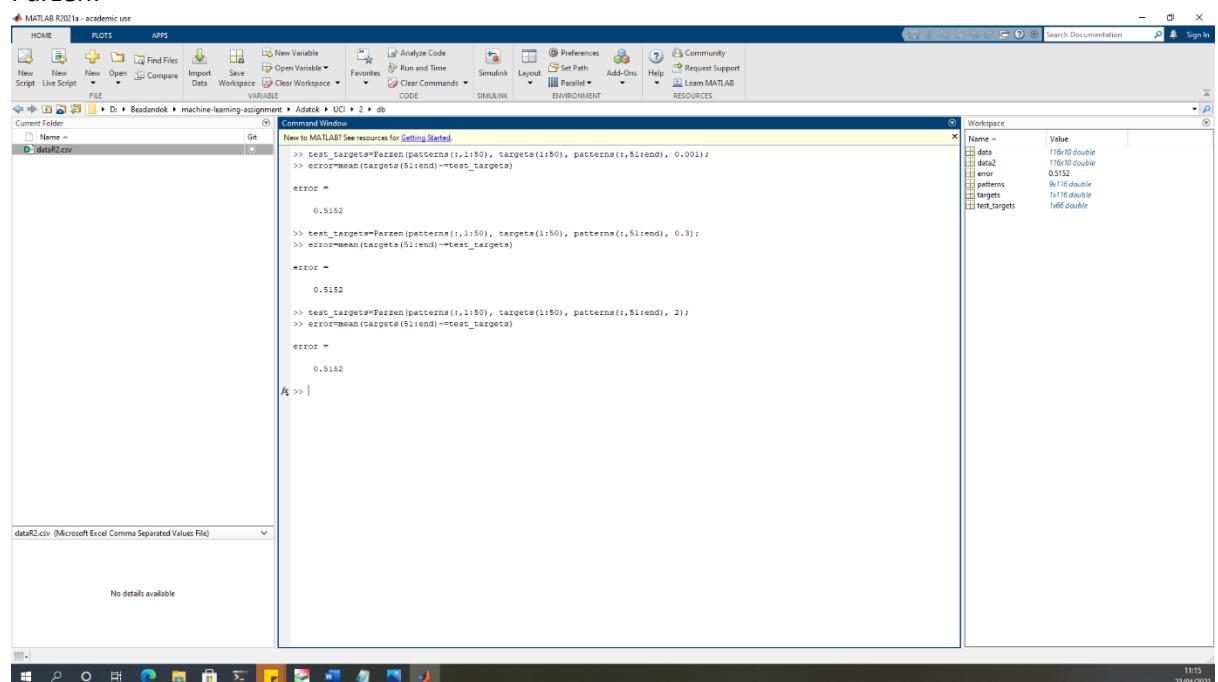
```
>> test_targets=Nearest_Neighbor(patterns(:,1:50), targets(1:50), patterns(:,51:end), 2);
>> error=mean(targets(51:end))-test_targets;
error =
0.4242
>> test_targets=Nearest_Neighbor(patterns(:,1:50), targets(1:50), patterns(:,51:end), 10);
>> error=mean(targets(51:end))-test_targets;
error =
0.4948
>> test_targets=Nearest_Neighbor(patterns(:,1:50), targets(1:50), patterns(:,51:end), 20);
>> error=mean(targets(51:end))-test_targets;
error =
0.5152
f1 >>
```

The Workspace browser on the right side shows the following variables:

Name	Type	Value
data	1x10 double	
data2	1x10 double	
error	0.4242	
patterns	9x16 double	
targets	1x16 double	
test_targets	1x10 double	

Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev

4. Parzen:



The screenshot shows the MATLAB R2021a interface. The Command Window displays the following code and its execution results:

```
>> test_targets=Parzen(patterns(:,1:50), targets(1:50), patterns(:,51:end), 0.001);
>> error=mean(targets(51:end))-test_targets

error =
0.5152

>> test_targets=Parzen(patterns(:,1:50), targets(1:50), patterns(:,51:end), 0.3);
>> error=mean(targets(51:end))-test_targets

error =
0.5152

>> test_targets=Parzen(patterns(:,1:50), targets(1:50), patterns(:,51:end), 2);
>> error=mean(targets(51:end))-test_targets

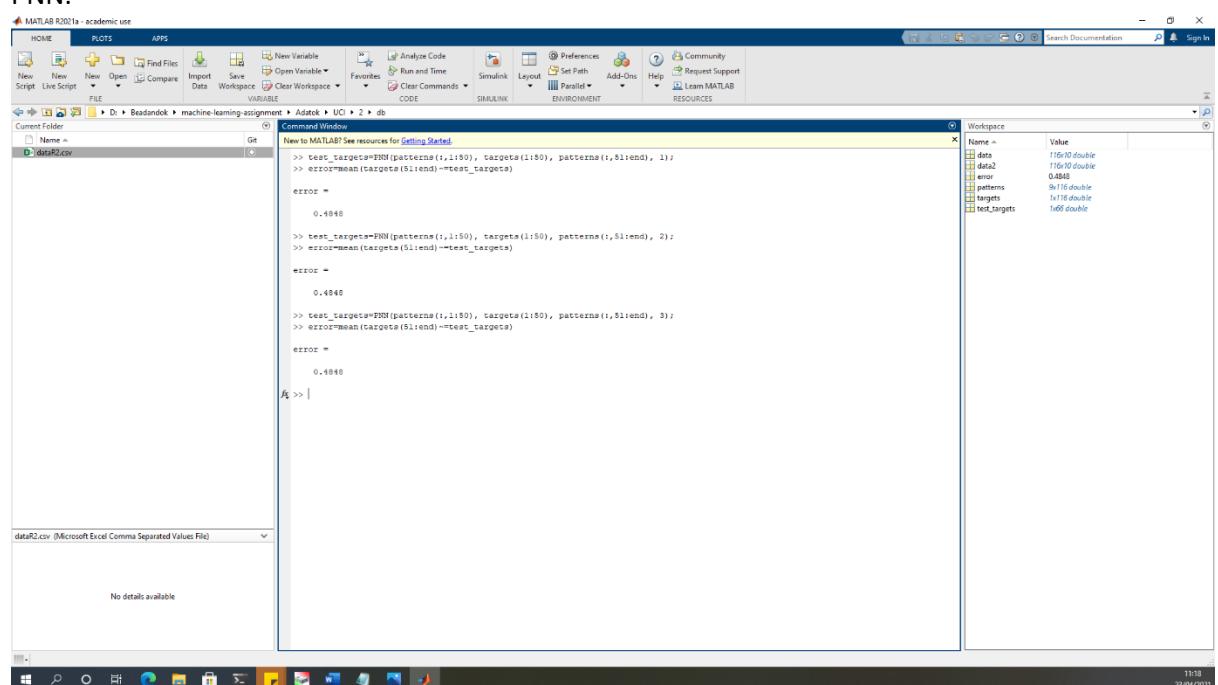
error =
0.5152

f2 >> |
```

The Workspace window on the right shows the variables defined:

Name	Type
data	1x10 double
data2	1x10 double
error	0.5152
patterns	9x16 double
targets	1x16 double
test_targets	1x10 double

5. PNN:



The screenshot shows the MATLAB R2021a interface. The Command Window displays the following code and its execution results:

```
>> test_targets=PNN(patterns(:,1:50), targets(1:50), patterns(:,51:end), 1);
>> error=mean(targets(51:end))-test_targets

error =
0.4848

>> test_targets=PNN(patterns(:,1:50), targets(1:50), patterns(:,51:end), 2);
>> error=mean(targets(51:end))-test_targets

error =
0.4848

>> test_targets=PNN(patterns(:,1:50), targets(1:50), patterns(:,51:end), 3);
>> error=mean(targets(51:end))-test_targets

error =
0.4848

f2 >> |
```

The Workspace window on the right shows the variables defined:

Name	Type
data	1x10 double
data2	1x10 double
error	0.4848
patterns	9x16 double
targets	1x16 double
test_targets	1x10 double

Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev

6. C4_5:

The screenshot shows the MATLAB R2021a interface. The Command Window displays MATLAB code for a decision tree classification, specifically using the C4.5 algorithm. The workspace contains variables: data (116x10 double), data2 (116x10 double), error (0.2424), patterns (9x16 double), targets (1x16 double), and test_targets (1x6 double). The current folder browser shows a file named 'dataR2.csv'.

```
>> test_targets=C4_5(patterns(:,1:50), targets(1:50), patterns(:,51:end), 2);
Building tree
Classify test samples using the tree
>> error=mean(targets(51:end))-test_targets

error =
0.3788

>> test_targets=C4_5(patterns(:,1:50), targets(1:50), patterns(:,51:end), 10);
Building tree
Classify test samples using the tree
>> error=mean(targets(51:end))-test_targets

error =
0.3465

>> test_targets=C4_5(patterns(:,1:50), targets(1:50), patterns(:,51:end), 20);
Building tree
Classify test samples using the tree
>> error=mean(targets(51:end))-test_targets

error =
0.2424
f2 >> |
```

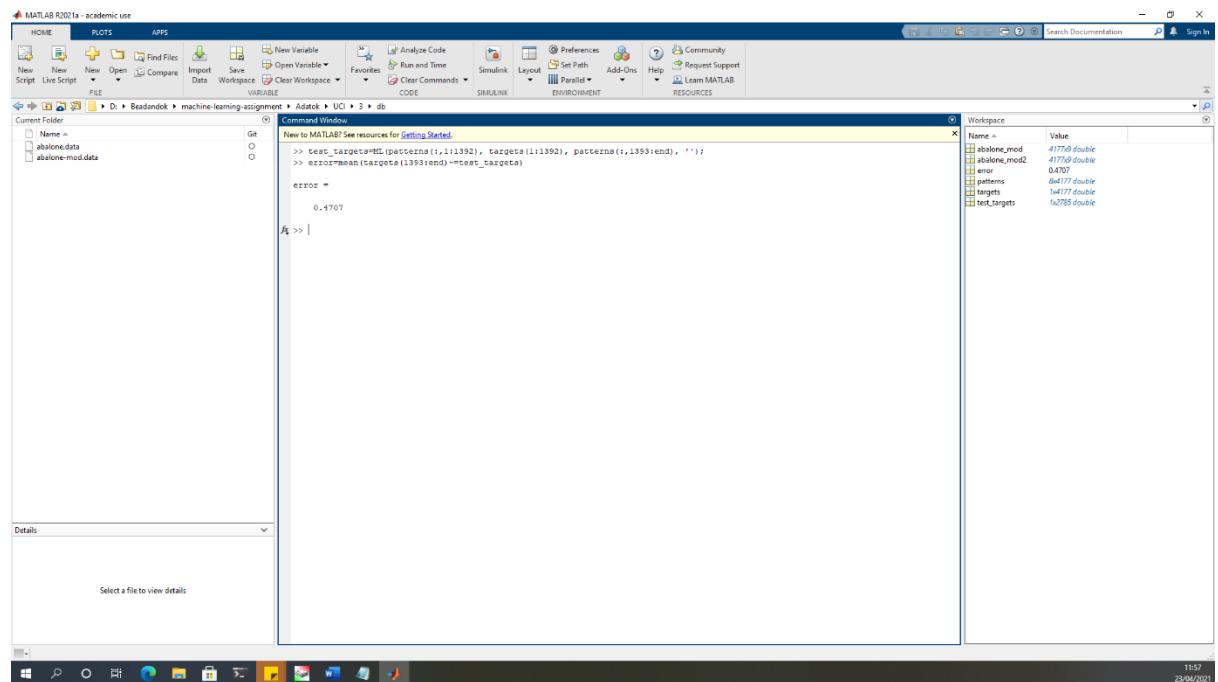
No details available

3. UCI adatbazis

Adatok	
Neve:	Abalone
Sorainak szama:	4177
Jellemzoinek szama:	8
Osztalyainak szama:	3
Elvegzett atalakitasok (ha szukseg es volt):	Lementettem a kozzettet data kiterjesztese fajlt (abalone.data), majd atalakitottam (abalone-mod.data). A cimkeosztalyok M, F, I ertekeit atalakitottam 1-3-ra, replace all-lal). Betoltottem a MATLAB-ba a tanult load parancsal.
Sikertelen futas oka, leirasa (ha volt ilyen):	Lent (ha volt ilyen)

A futasrol készült kepernyokepek (Rajtuk a hibaszazalek)

1. ML:



2. ML_diag:

The screenshot shows the MATLAB R2020a interface. The Command Window displays the following code and output:

```
>> test_targets=ML_diag(patterns(:,1:1392), targets(1:1392), patterns(:,1393:end), 1);
>> error=mean(targets(1393:end)-test_targets)

error =
0.4948

f1 >>
```

The Workspace browser on the right shows the following variables:

Name	Type
abalone_mod	41779 double
abalone_mod2	41779 double
error	0.4948
patterns	4x177 double
targets	1x177 double
test_targets	1x2785 double

3. Nearest_neighbor:

The screenshot shows the MATLAB R2020a interface. The Command Window displays the following code and output:

```
>> test_targets=Nearest_Neighbor(patterns(:,1:1392), targets(1:1392), patterns(:,1393:end), 2);
>> error=mean(targets(1393:end)-test_targets)

error =
0.5189

>> test_targets=Nearest_Neighbor(patterns(:,1:1392), targets(1:1392), patterns(:,1393:end), 10);
>> error=mean(targets(1393:end)-test_targets)

error =
0.4740

>> test_targets=Nearest_Neighbor(patterns(:,1:1392), targets(1:1392), patterns(:,1393:end), 20);
>> error=mean(targets(1393:end)-test_targets)

error =
0.45978

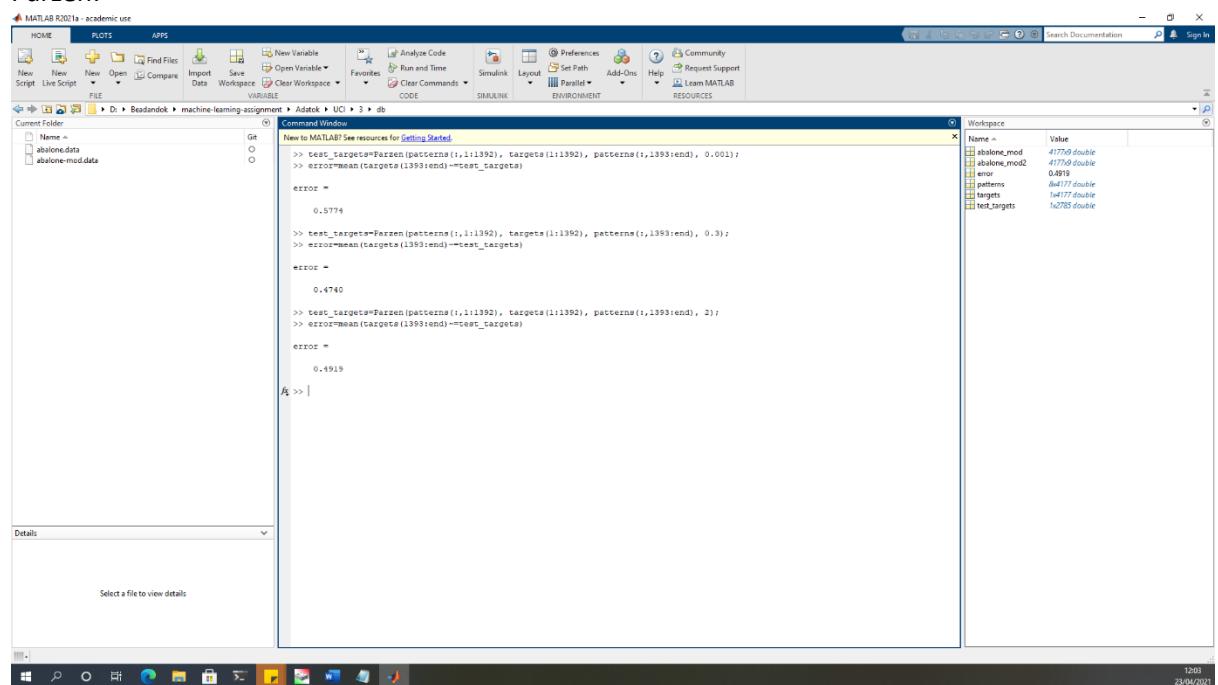
f1 >>
```

The Workspace browser on the right shows the following variables:

Name	Type
abalone_mod	41779 double
abalone_mod2	41779 double
error	0.45978
patterns	4x177 double
targets	1x177 double
test_targets	1x2785 double

Keszítette: Bozsik Armand Viktor
 Programtervező Informatikus MSc (levelező) szakos hallgató
 2020/21 II. felev

4. Parzen:



The screenshot shows the MATLAB interface with the following details:

- Command Window:** Displays the command history for training a Parzen classifier. The code includes loading data, defining target variables, and running the `Parzen` function three times with different parameters (1, 2, and 3). The error values are printed after each run.
- Workspace Browser:** Shows the variables defined in the workspace, including `abalone_mod`, `error`, `patterns`, `targets`, and `test_targets`.
- Taskbar:** Shows the Windows taskbar with various application icons.

```

>> test_targets=Parzen(patterns(1:11392), targets(1:11392), patterns(1,1393:end), 0.001);
>> error=mean(targets(1393:end))-test_targets
error =
0.5774

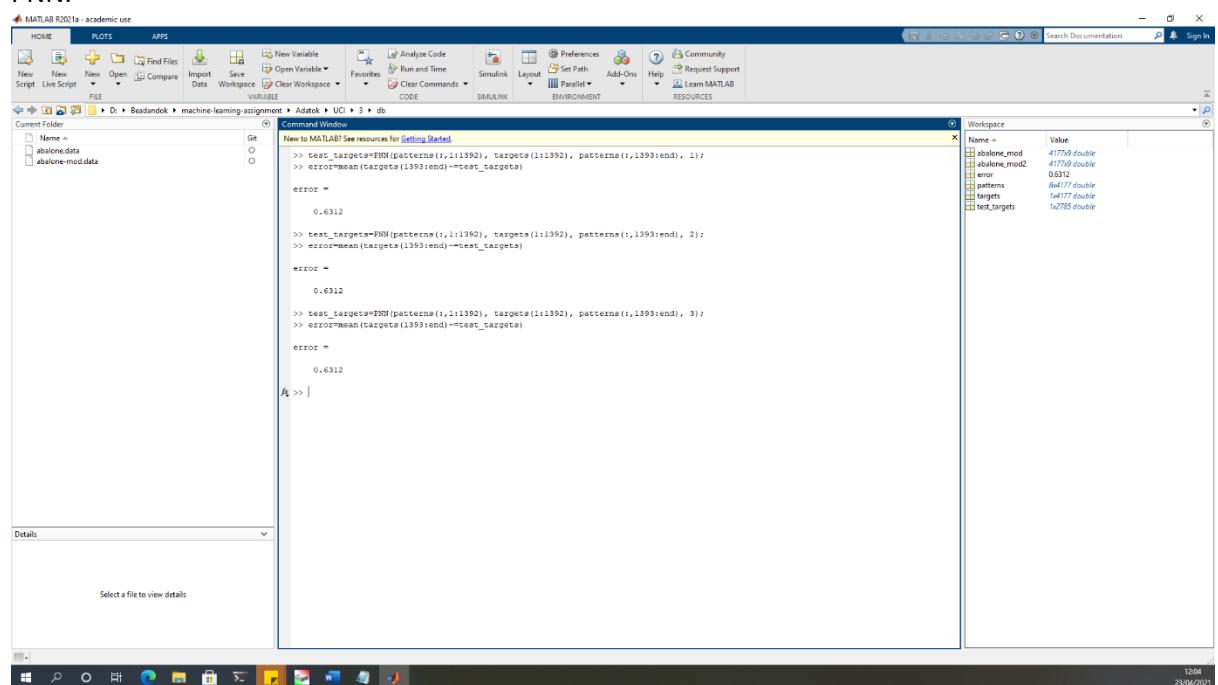
>> test_targets=Parzen(patterns(1:11392), targets(1:11392), patterns(1,1393:end), 0.3);
>> error=mean(targets(1393:end))-test_targets
error =
0.4740

>> test_targets=Parzen(patterns(1:11392), targets(1:11392), patterns(1,1393:end), 2);
>> error=mean(targets(1393:end))-test_targets
error =
0.4919

>> >

```

5. PNN:



The screenshot shows the MATLAB interface with the following details:

- Command Window:** Displays the command history for training a PNN classifier. The code includes loading data, defining target variables, and running the `PNN` function three times with different parameters (1, 2, and 3). The error values are printed after each run.
- Workspace Browser:** Shows the variables defined in the workspace, including `abalone_mod`, `abalone_mod2`, `error`, `patterns`, `targets`, and `test_targets`.
- Taskbar:** Shows the Windows taskbar with various application icons.

```

>> test_targets=PNN(patterns(1:11392), targets(1:11392), patterns(1,1393:end), 1);
>> error=mean(targets(1393:end))-test_targets
error =
0.6312

>> test_targets=PNN(patterns(1:11392), targets(1:11392), patterns(1,1393:end), 2);
>> error=mean(targets(1393:end))-test_targets
error =
0.6312

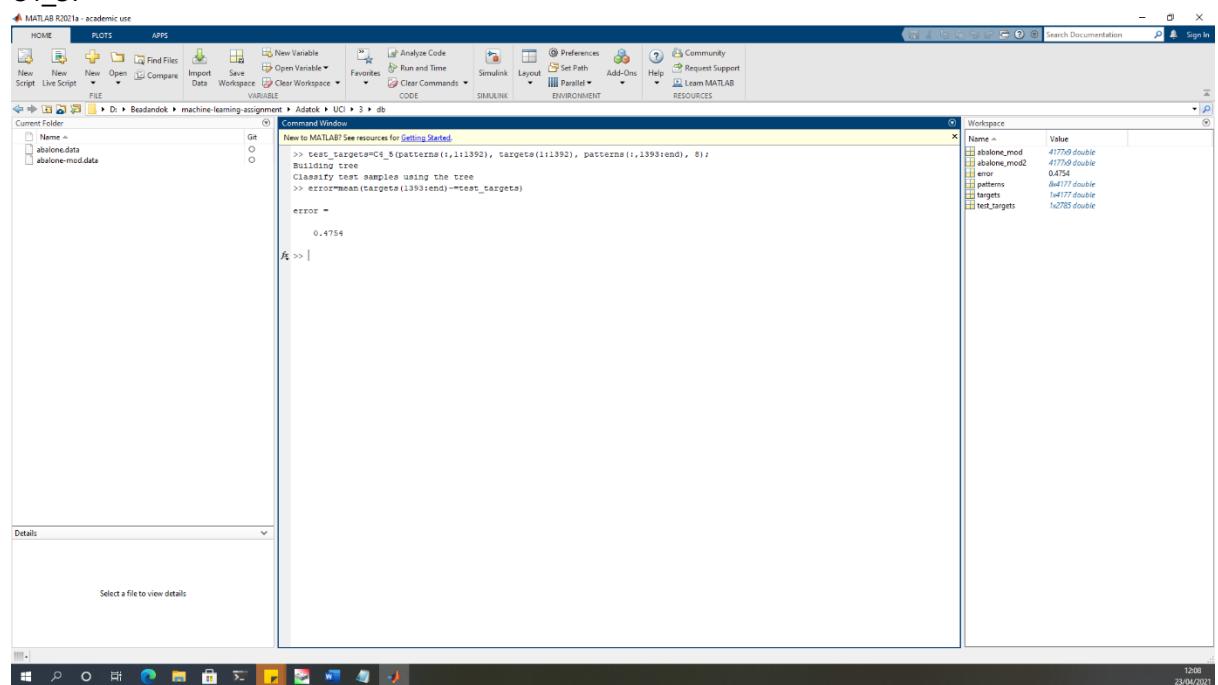
>> test_targets=PNN(patterns(1:11392), targets(1:11392), patterns(1,1393:end), 3);
>> error=mean(targets(1393:end))-test_targets
error =
0.6312

>> >

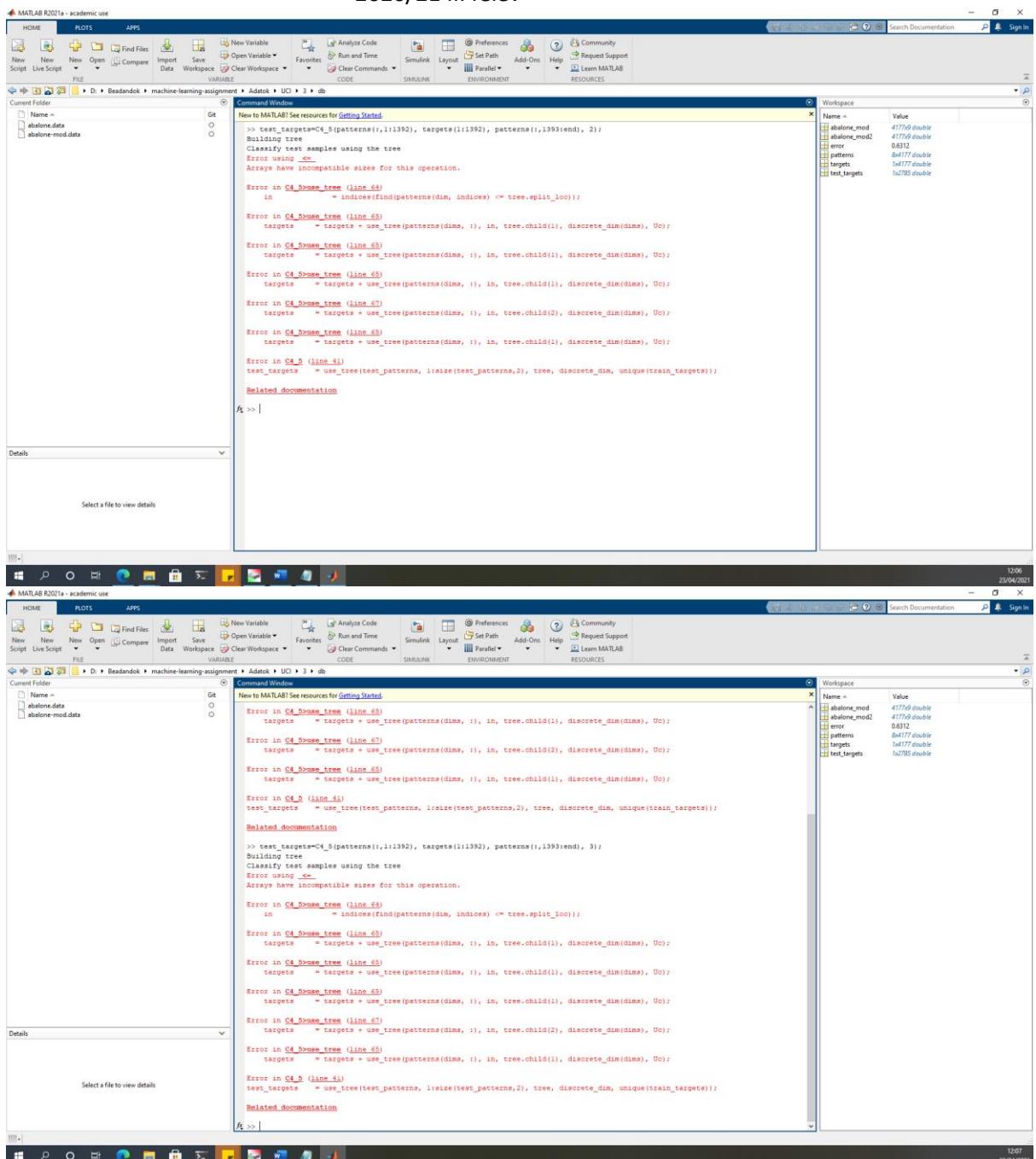
```

Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev

6. C4_5:



Keszítette: Bozsik Armand Viktor
 Programtervező Informatikus MSc (levelező) szakos hallgató
 2020/21 II. felev



```

>> test_targets=C4_5(patterns(:,1:1392), targets(:,1:1392), patterns(:,1393:end), 2);
Building tree
Classify test samples using the tree
Error using <math>\_c4_5</math>
Arrays have incompatible sizes for this operation.

Error in <math>\_c4\_5</math>_tree (Line 64)
    in
        = indices(find(patterns(dim, indices) <= tree.split_loc));

Error in <math>\_c4\_5</math>_tree (Line 65)
    targets
        = targets + use_tree(patterns(dim, :), in, tree.child(1), discrete_dim(dim), Uc);

Error in <math>\_c4\_5</math>_tree (Line 66)
    targets
        = targets + use_tree(patterns(dim, :), in, tree.child(1), discrete_dim(dim), Uc);

Error in <math>\_c4\_5</math>_tree (Line 67)
    targets
        = targets + use_tree(patterns(dim, :), in, tree.child(1), discrete_dim(dim), Uc);

Error in <math>\_c4\_5</math>_tree (Line 68)
    targets
        = targets + use_tree(patterns(dim, :), in, tree.child(2), discrete_dim(dim), Uc);

Error in <math>\_c4\_5</math>_tree (Line 69)
    targets
        = targets + use_tree(patterns(dim, :), in, tree.child(1), discrete_dim(dim), Uc);

Error in <math>\_c4\_5</math>_tree (Line 70)
    targets
        = targets + use_tree(patterns(dim, :), in, tree.child(1), discrete_dim(dim), Uc);

Error in <math>\_c4\_5</math>_tree (Line 71)
    targets
        = use_tree(test_patterns, 1:size(test_patterns,2), tree, discrete_dim, unique(train_targets));

Related documentation

f2 >> |
```

Details

Select a file to view details

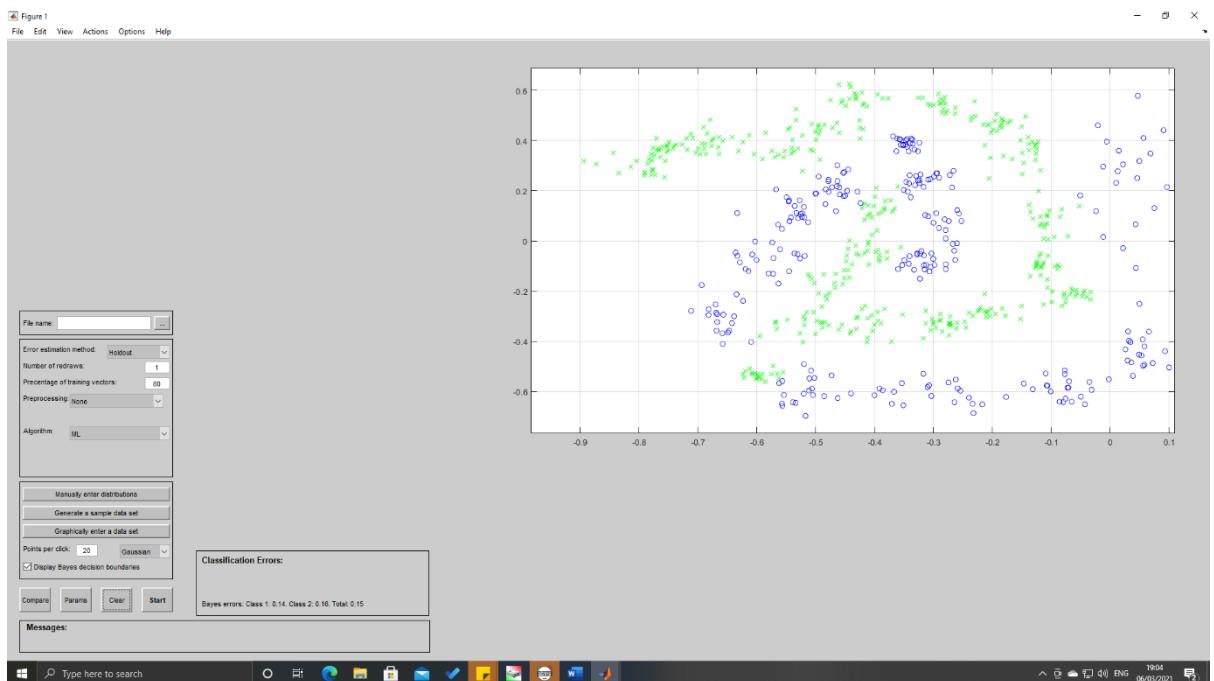
Name	Value
abalone_mod	41776 double
abalone_mod2	41776 double
error	0.6112
patterns	1x1777 double
targets	1x1777 double
test_targets	1x2785 double

Keszítette: Bozsik Armand Viktor
Programtervező Informatikus MSc (levelező) szakos hallgató
2020/21 II. felev
Saját adatbazisok

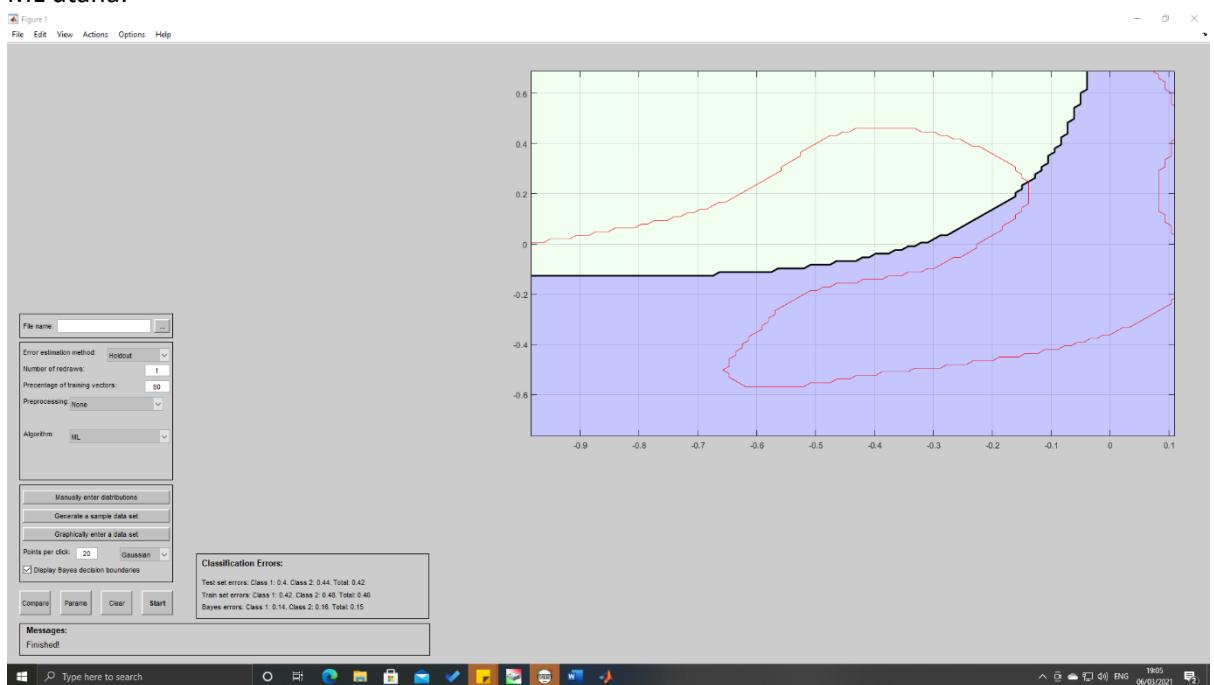
1. Saját adatbazis

A futásról készült kepernyőképek (Rajtuk a hibaszázalek)

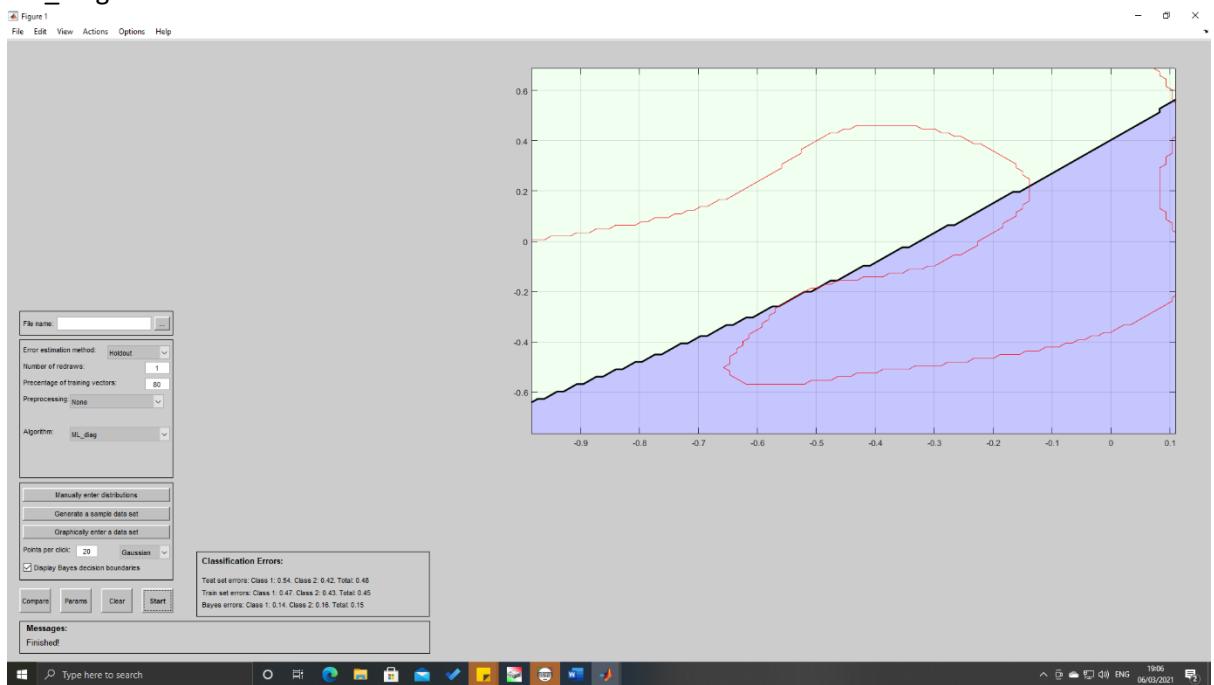
1. ML előtte:



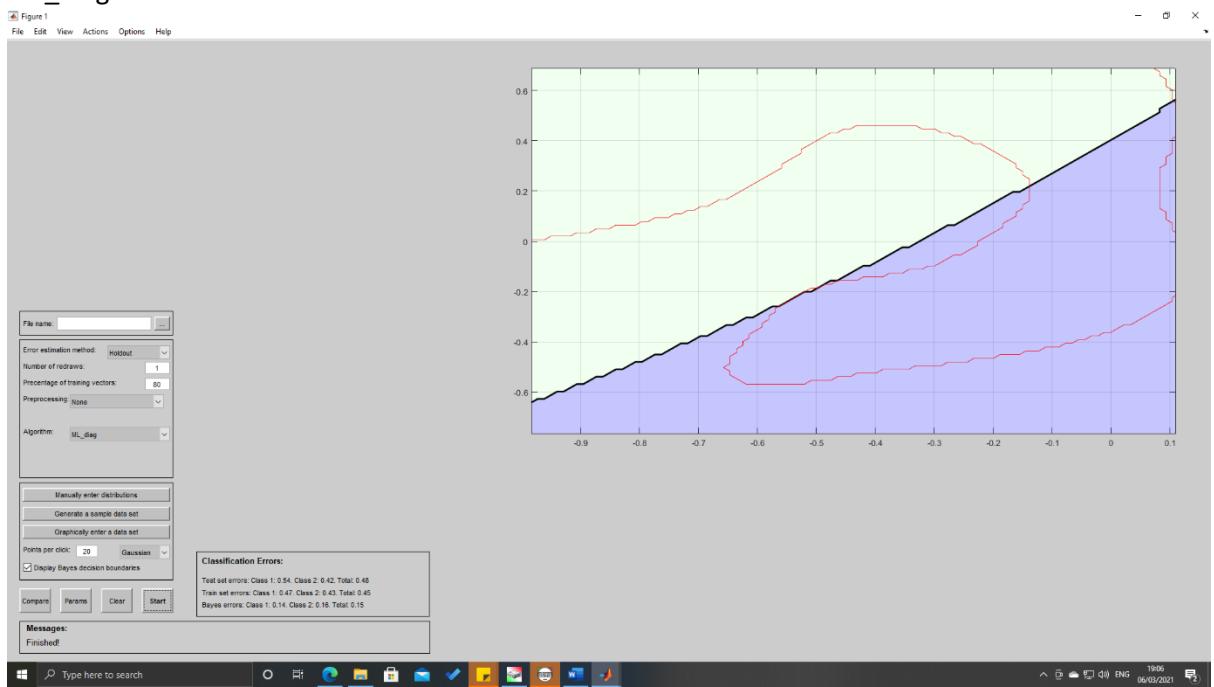
ML utána:



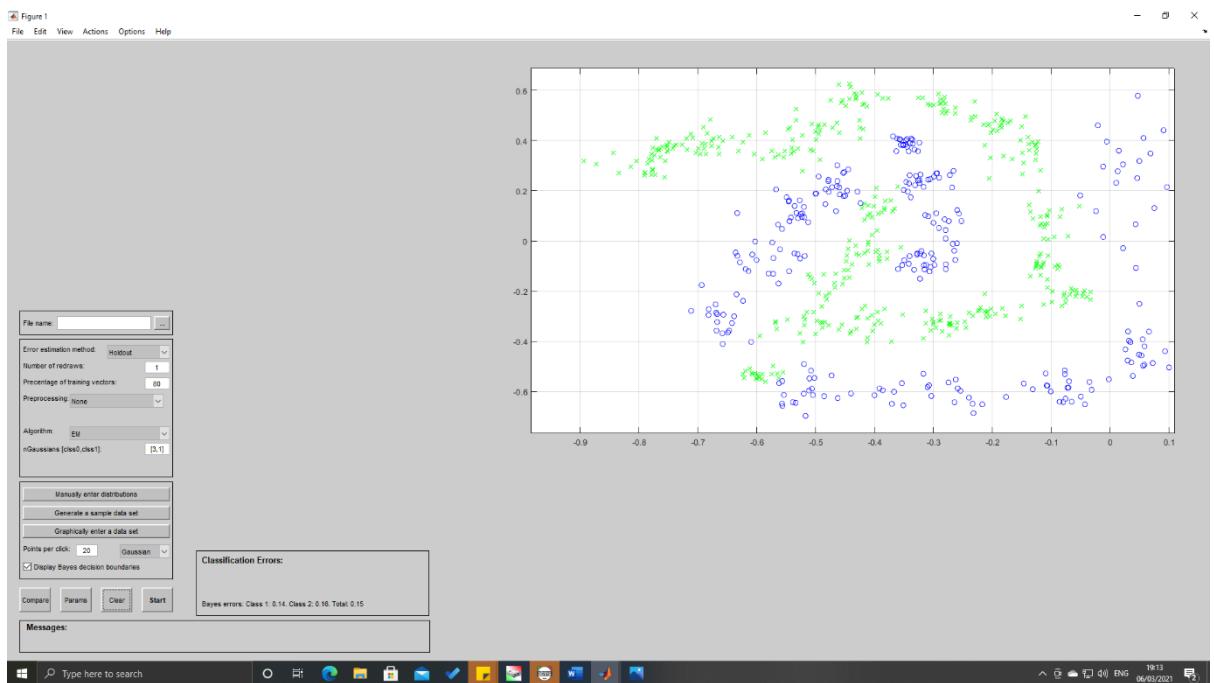
2. ML_diag előtte:



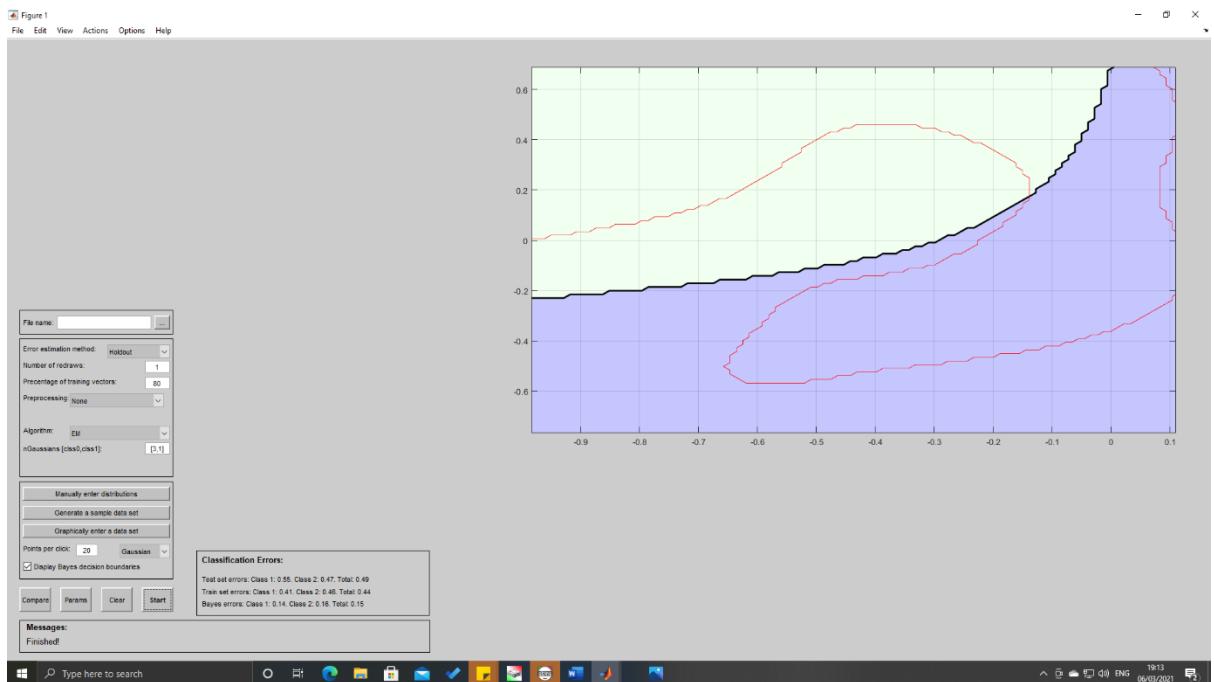
ML_diag utána:



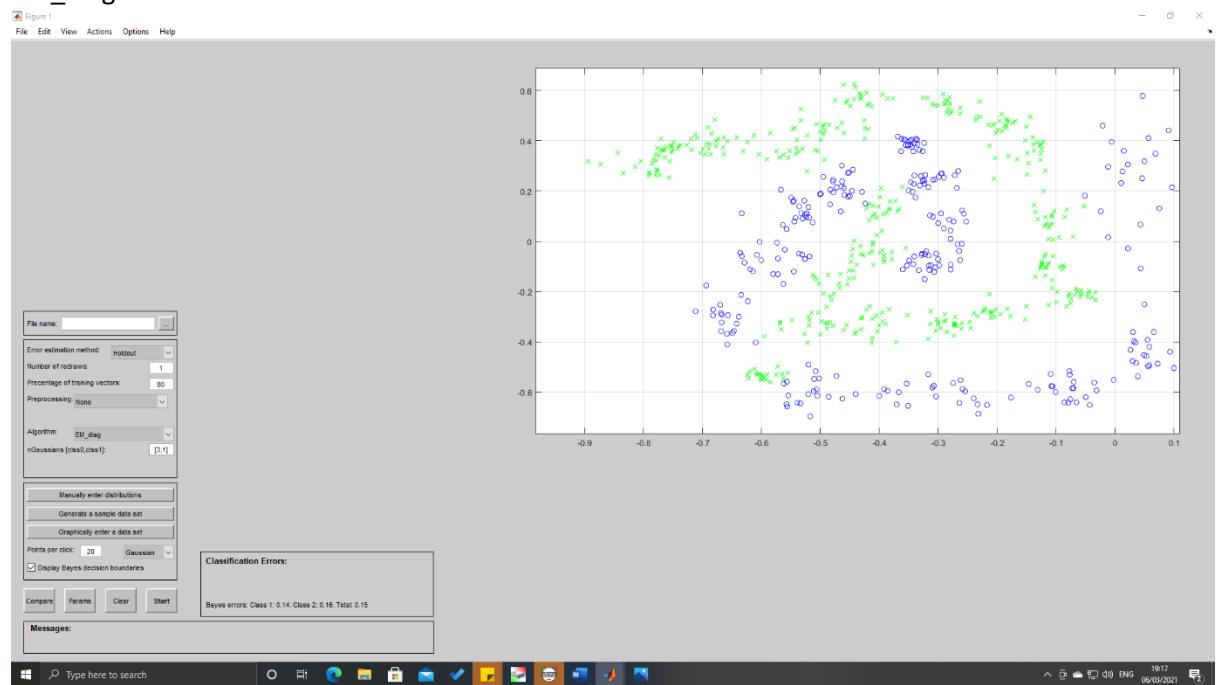
3. EM előtte:



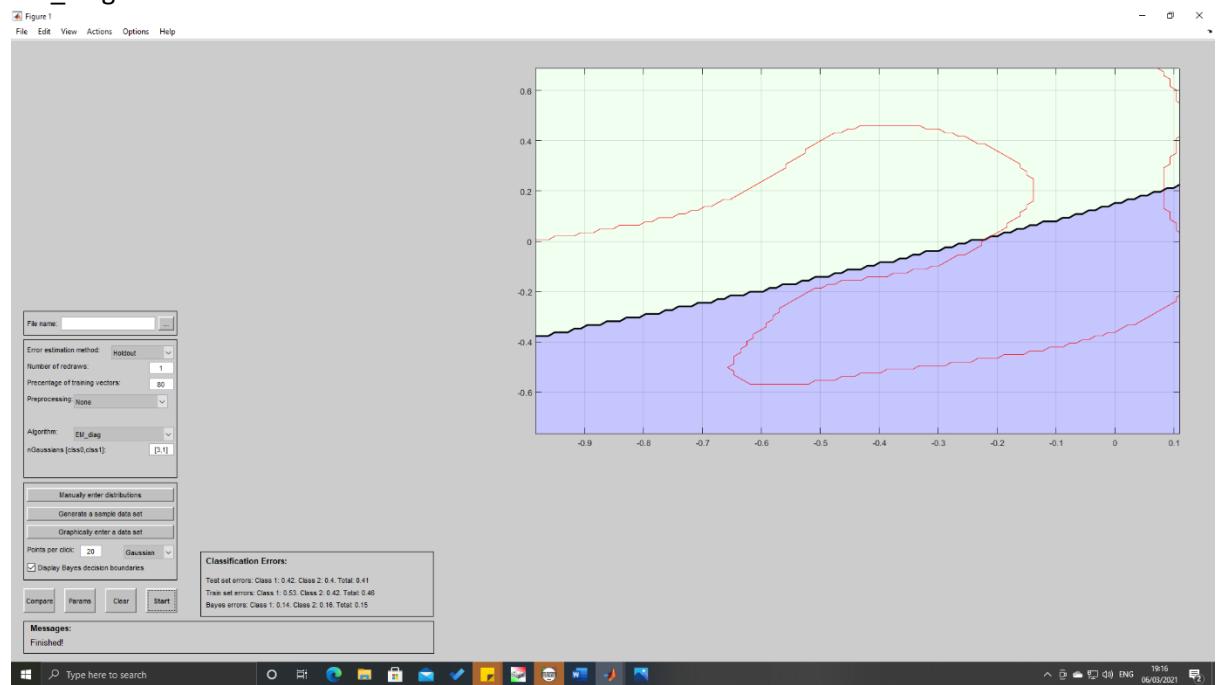
EM utána:



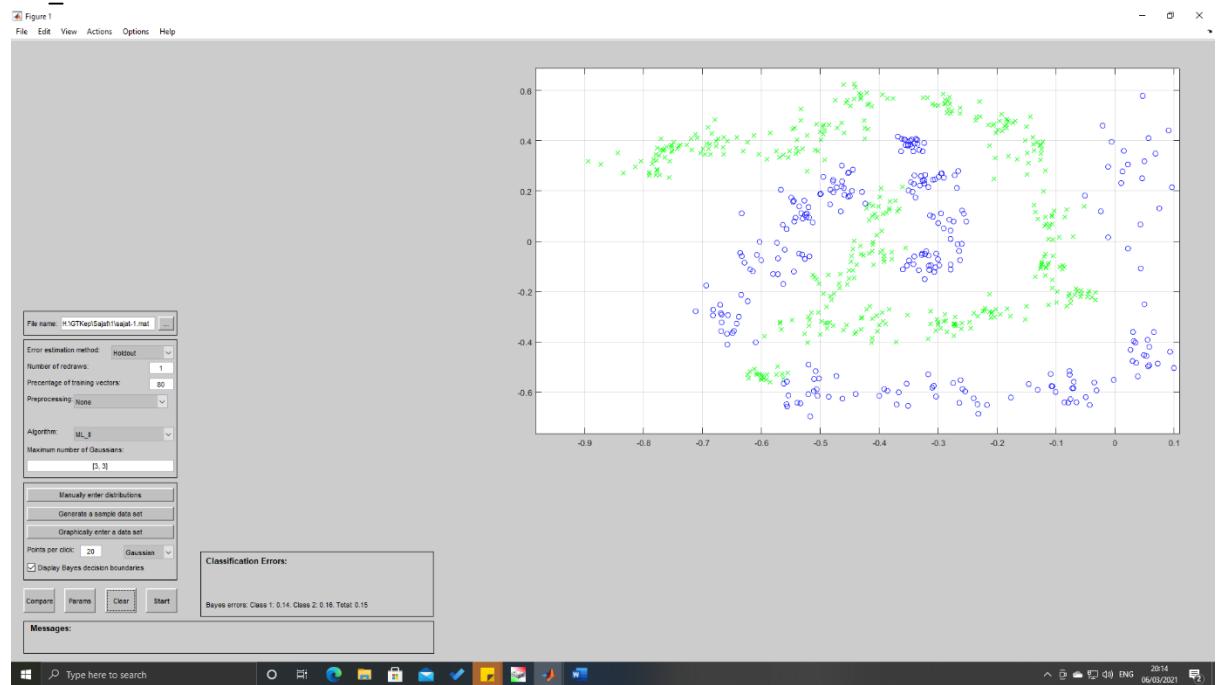
4. EM_diag előtte:



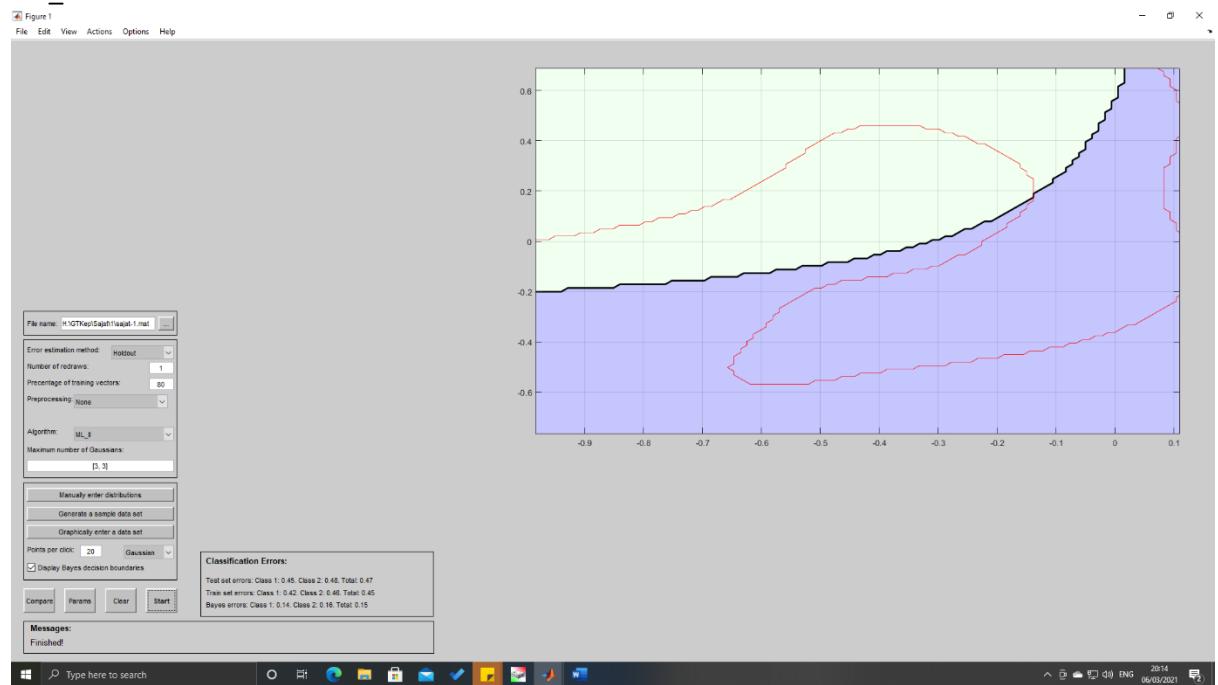
EM_diag utána:



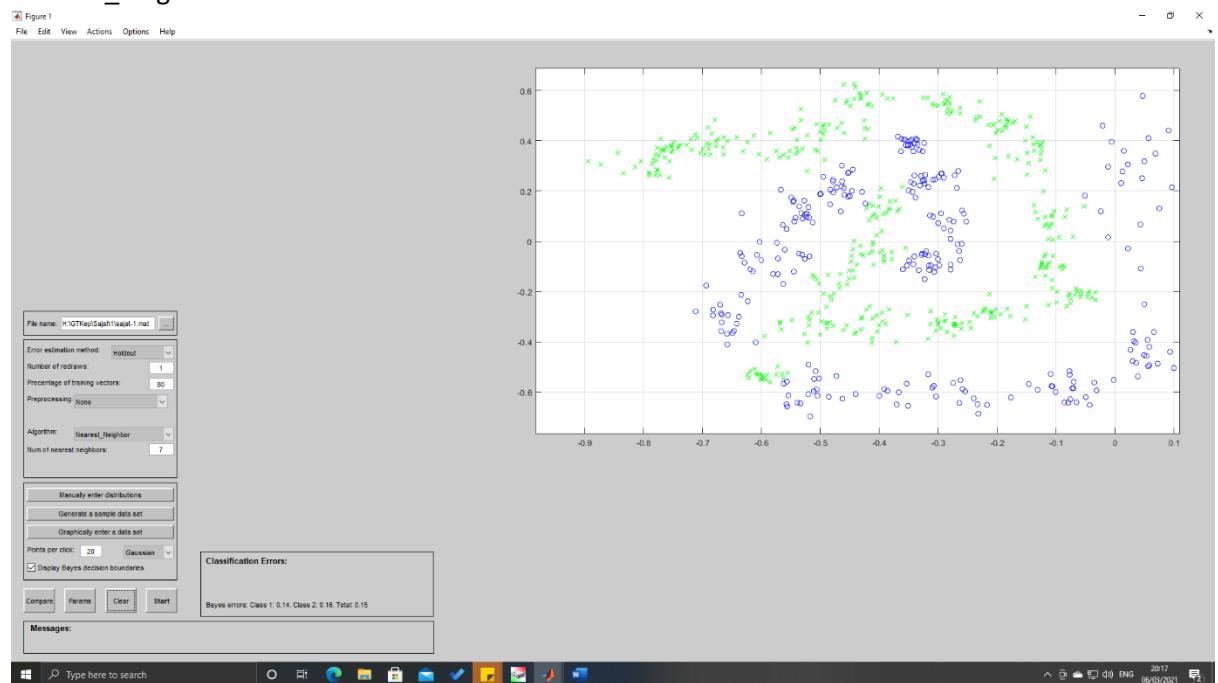
5. ML_II előtte:



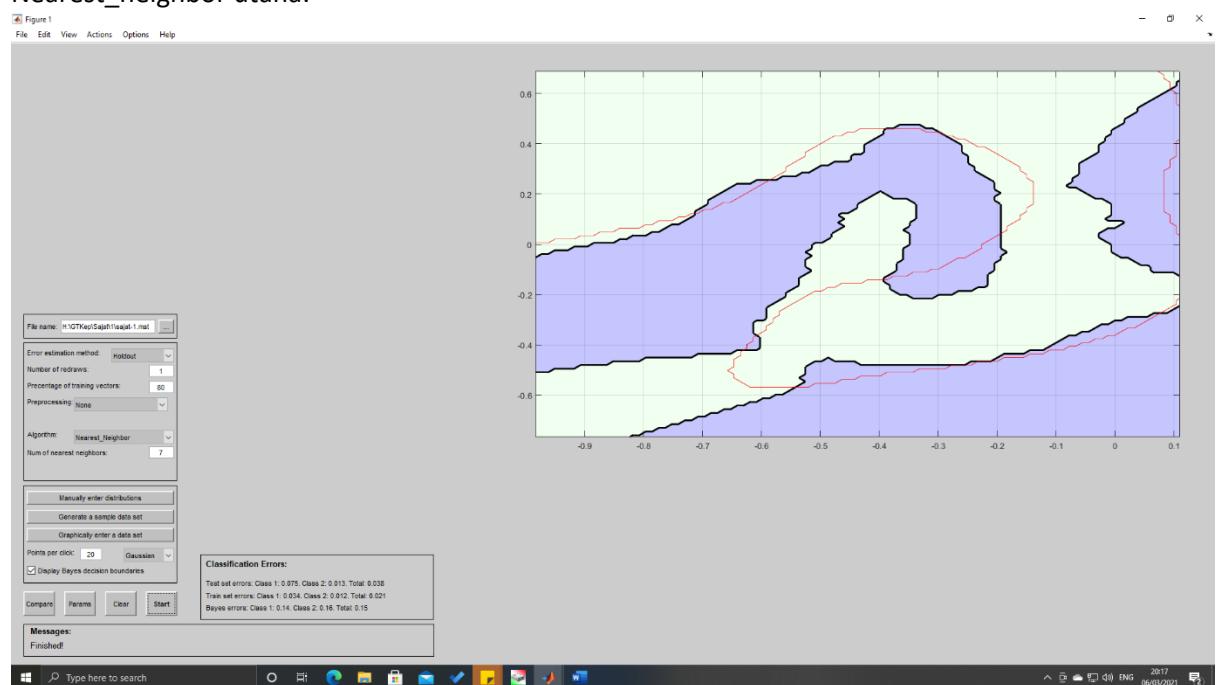
ML_II utána:



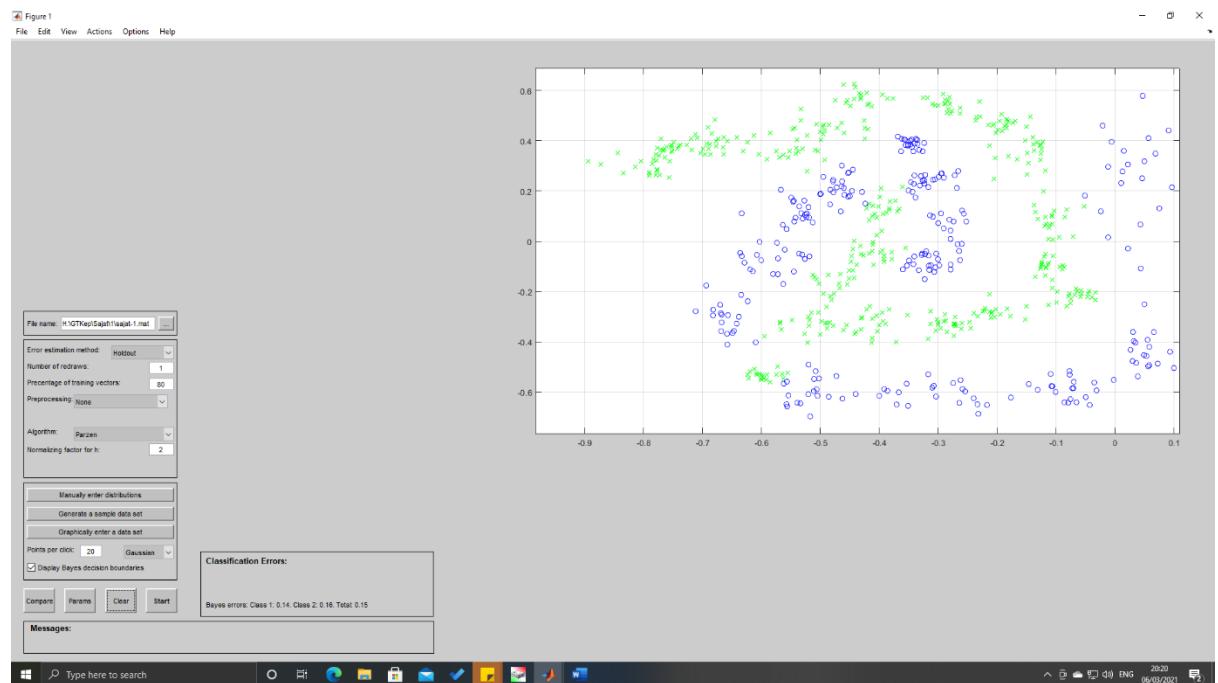
6. Nearest_neighbor elotte:



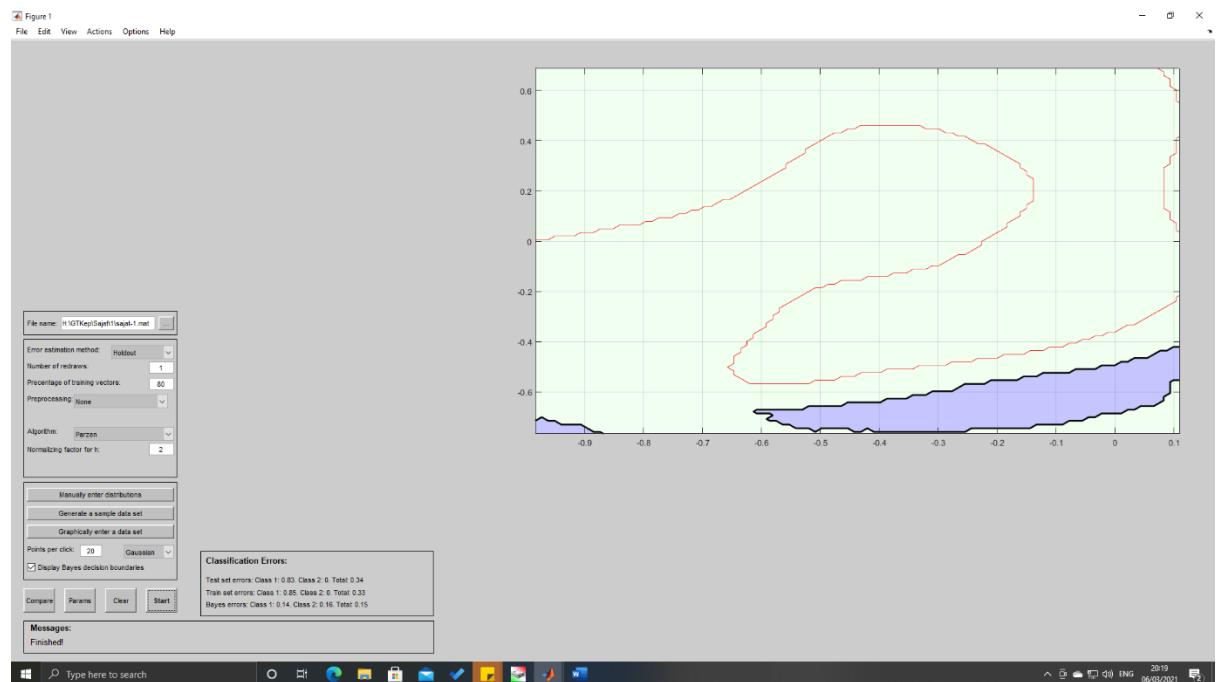
Nearest_neighbor utána:



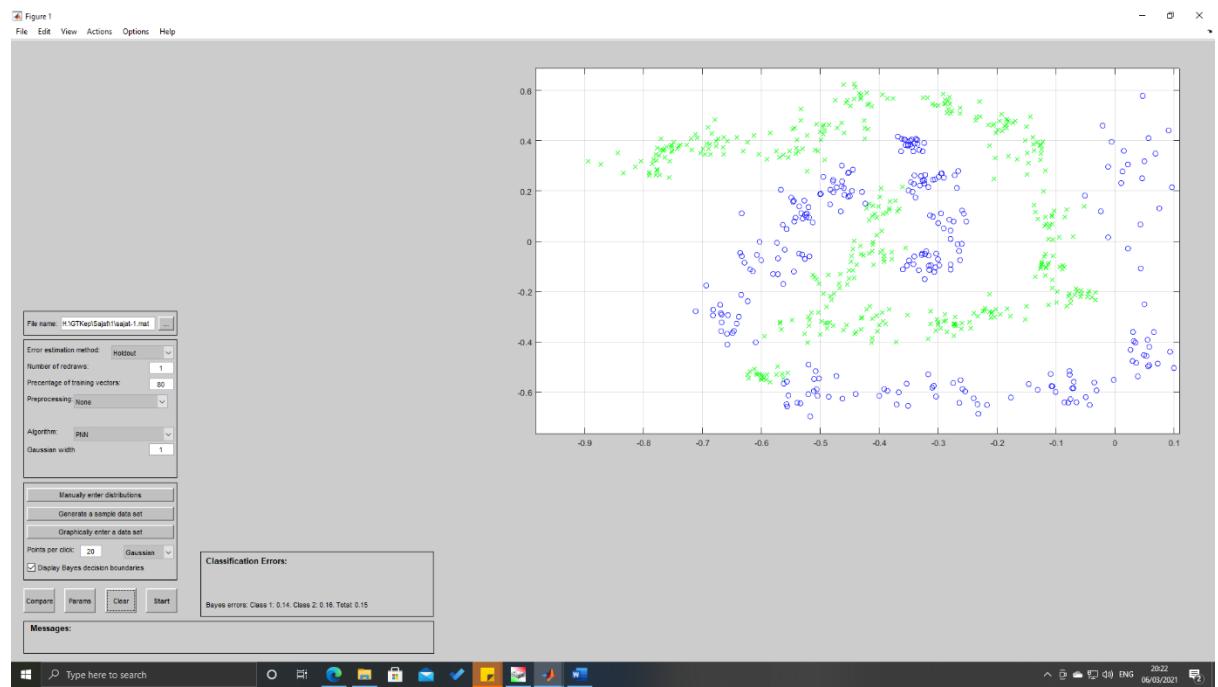
7. Parzen elotte:



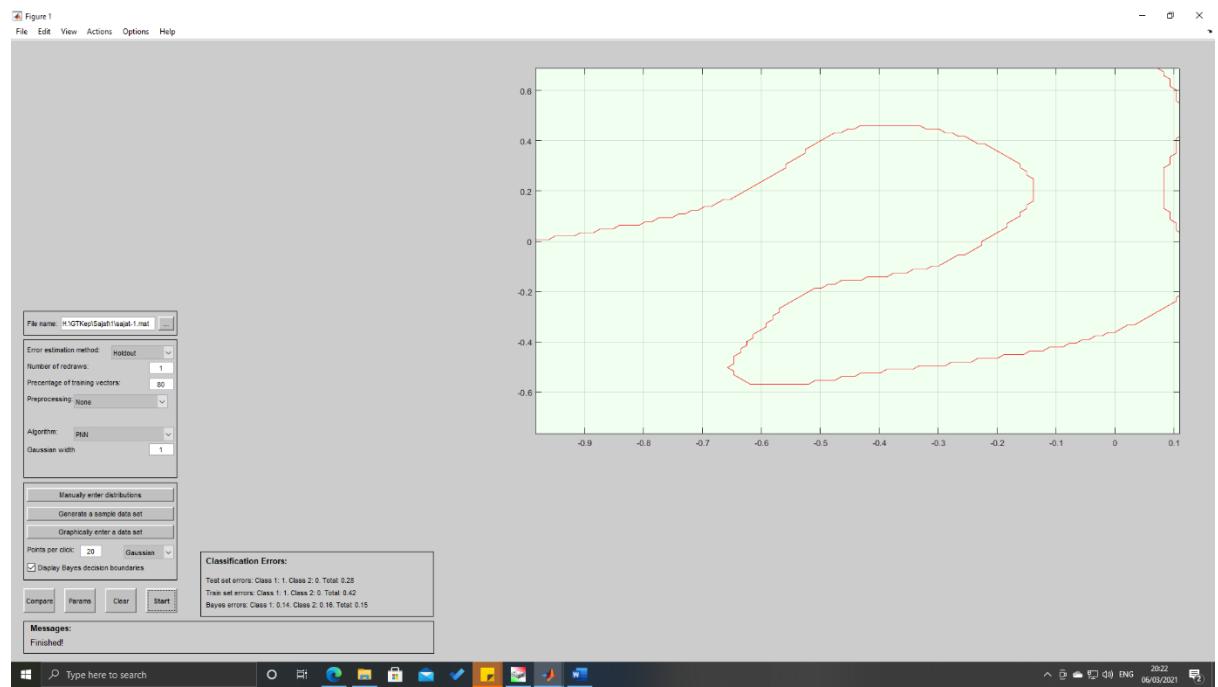
Parzen utana:



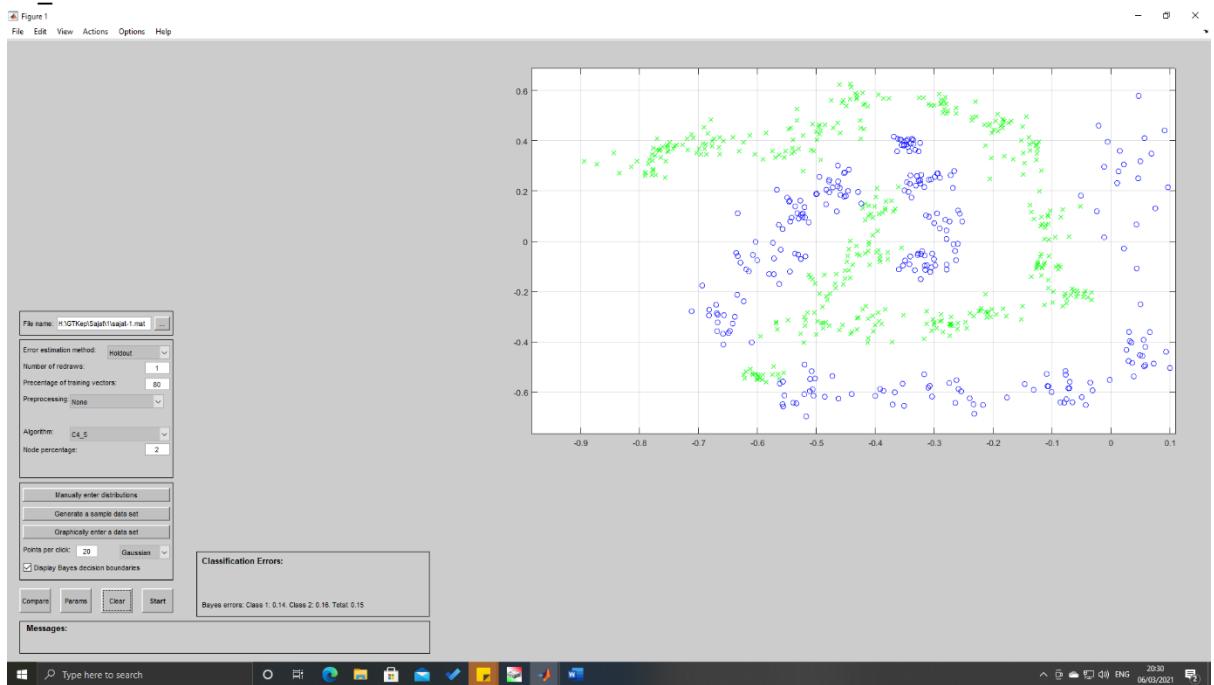
8. PNN előtte:



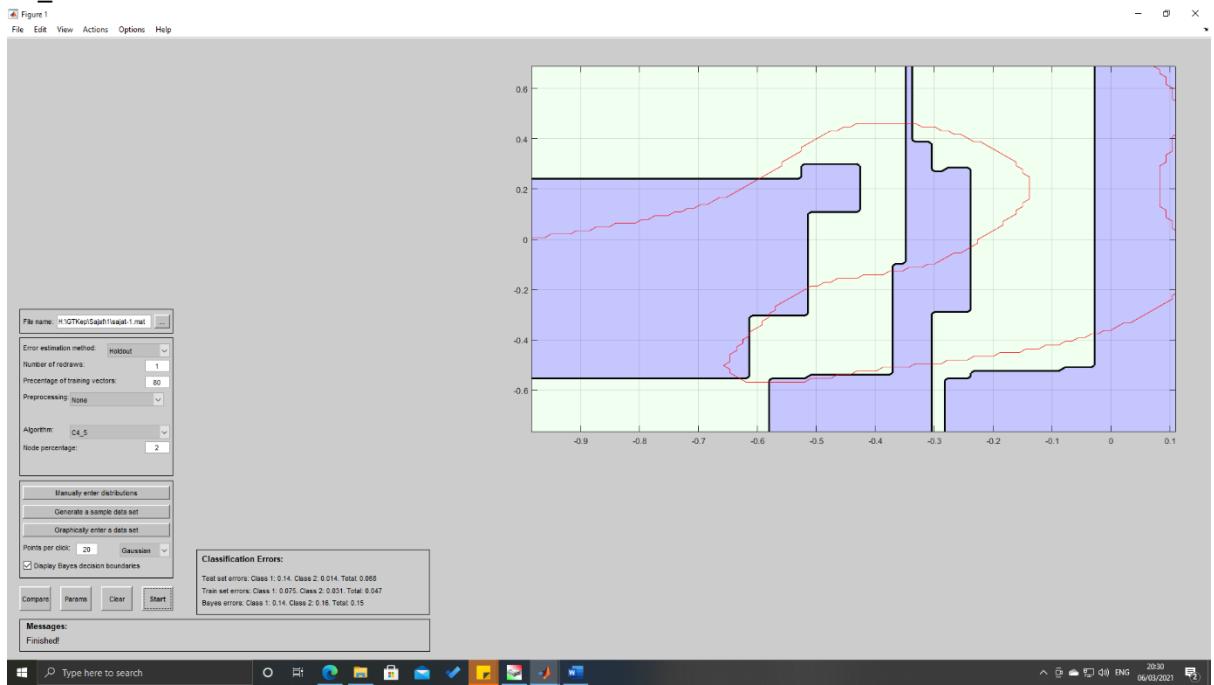
PNN utána:



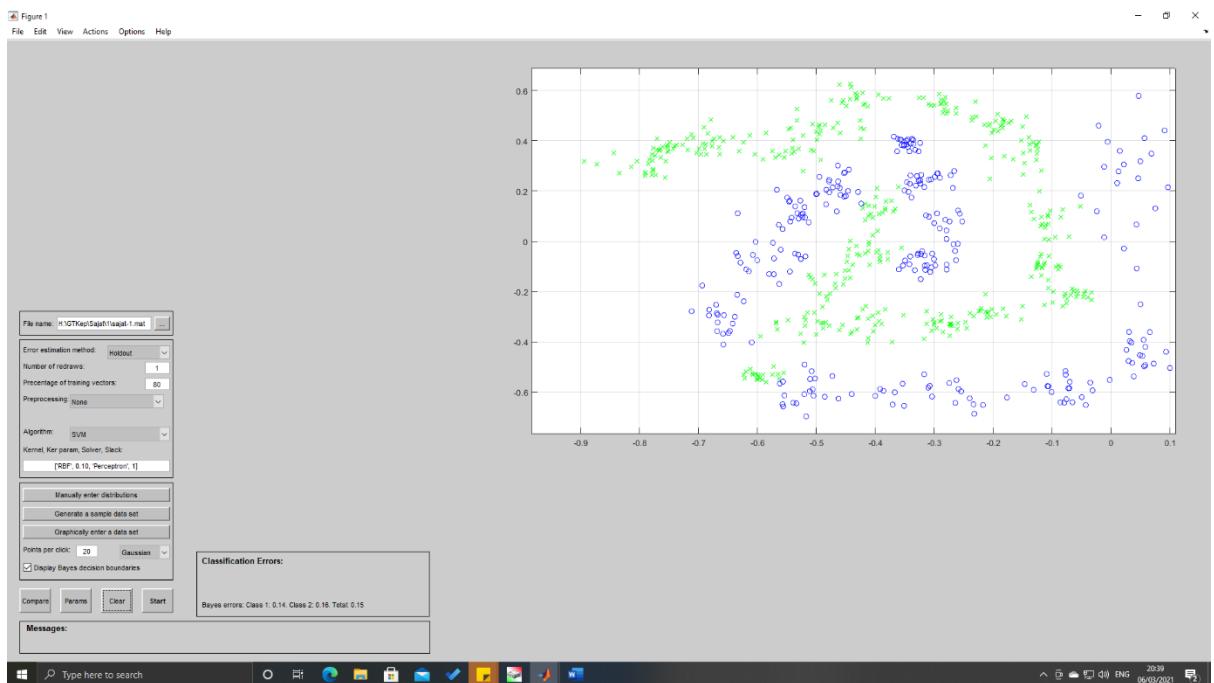
9. C4_5 elotte:



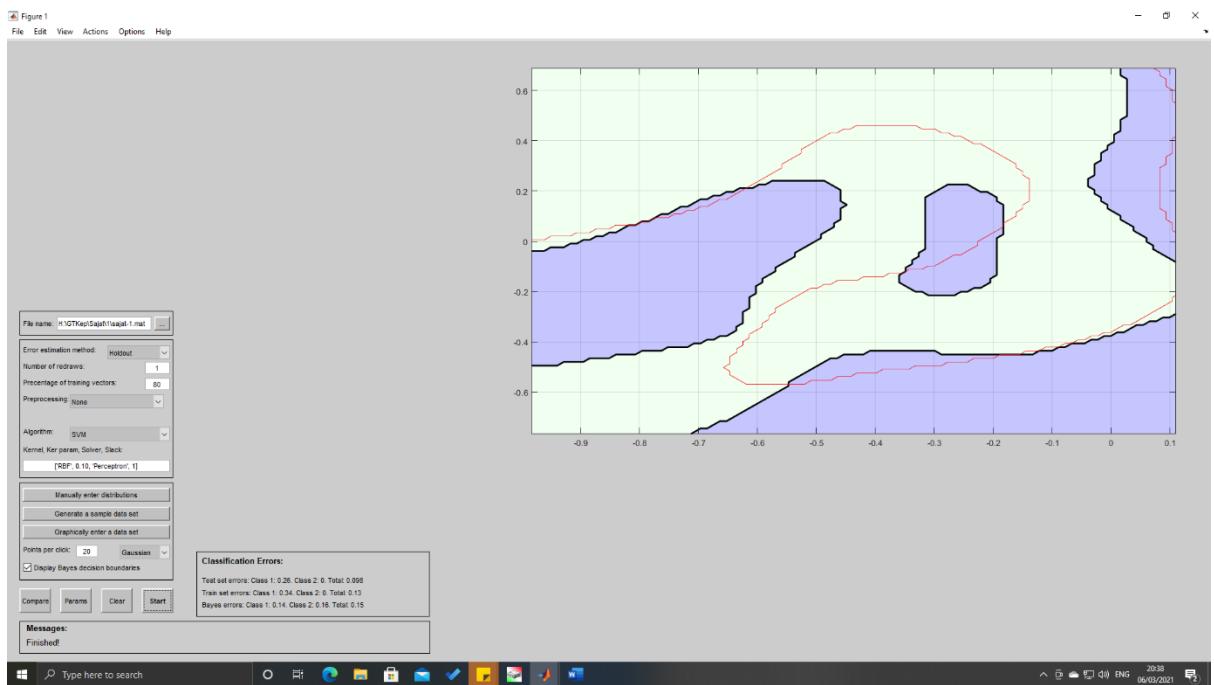
C4_5 utana:



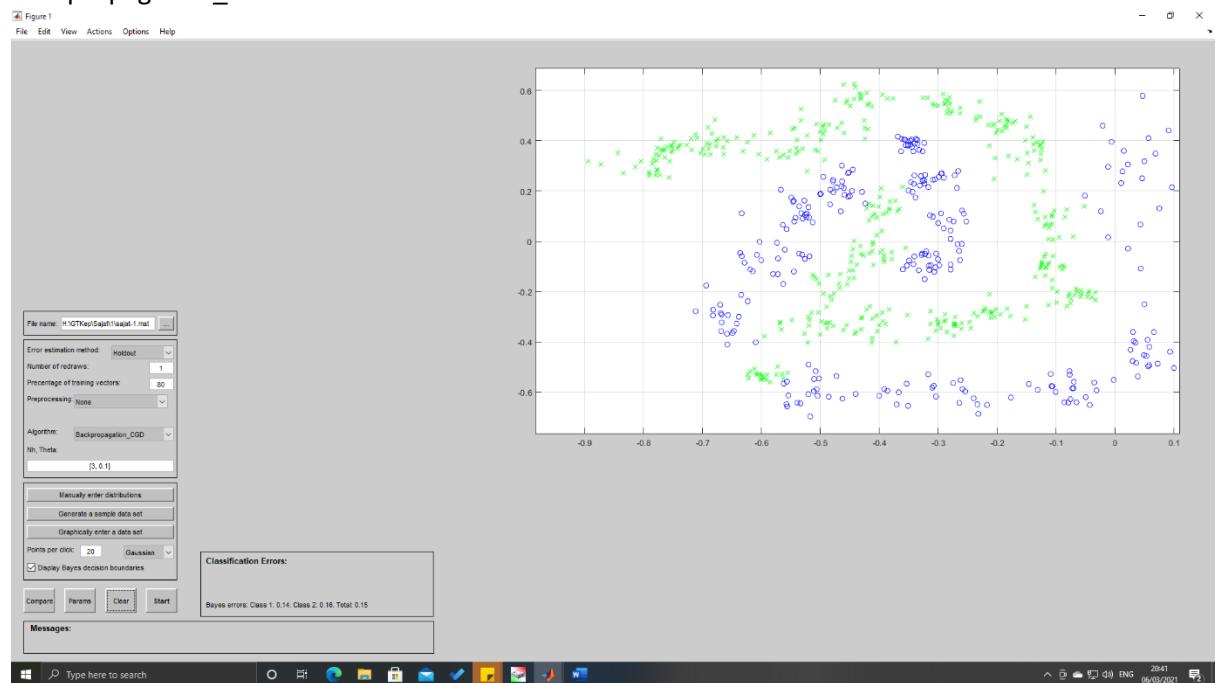
10. SVM előtte:



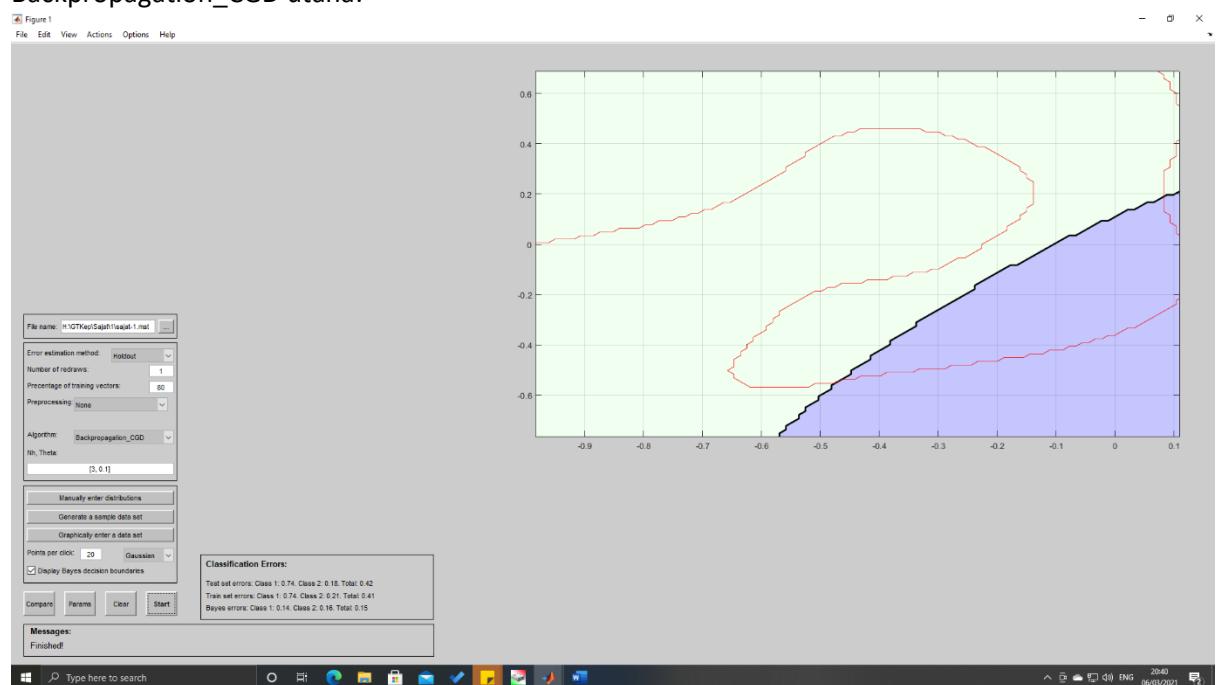
SVM utána:



11. Backpropagation_CGD előtte:



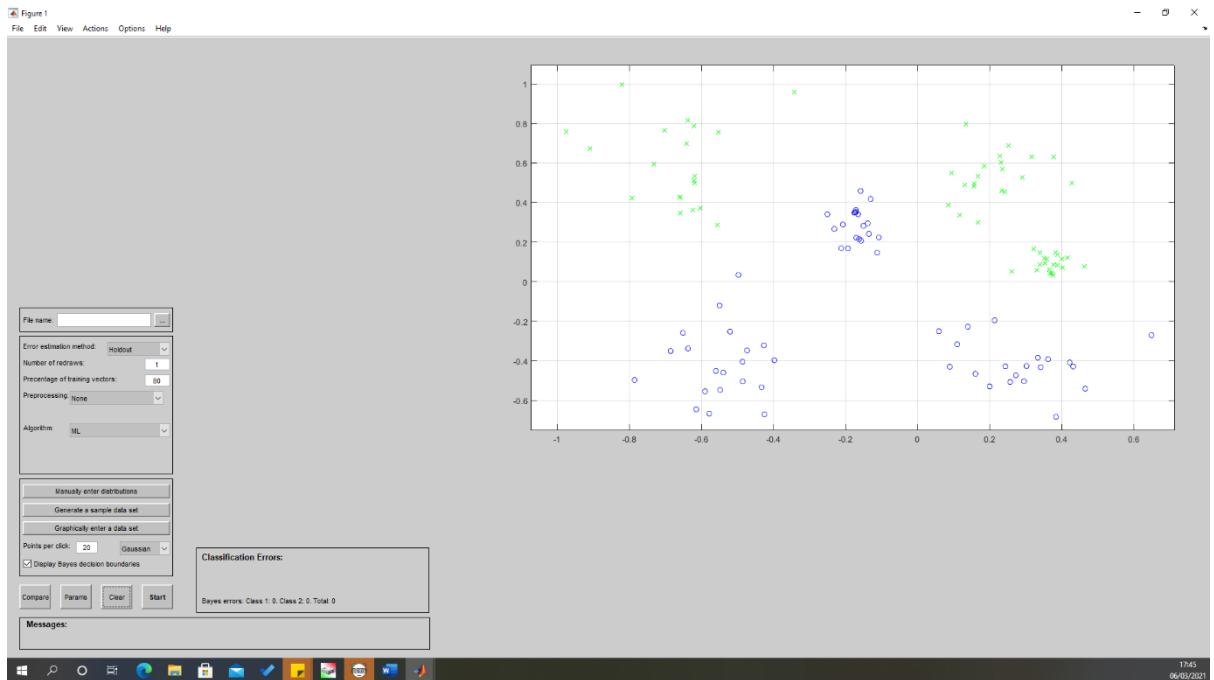
Backpropagation_CGD utána:



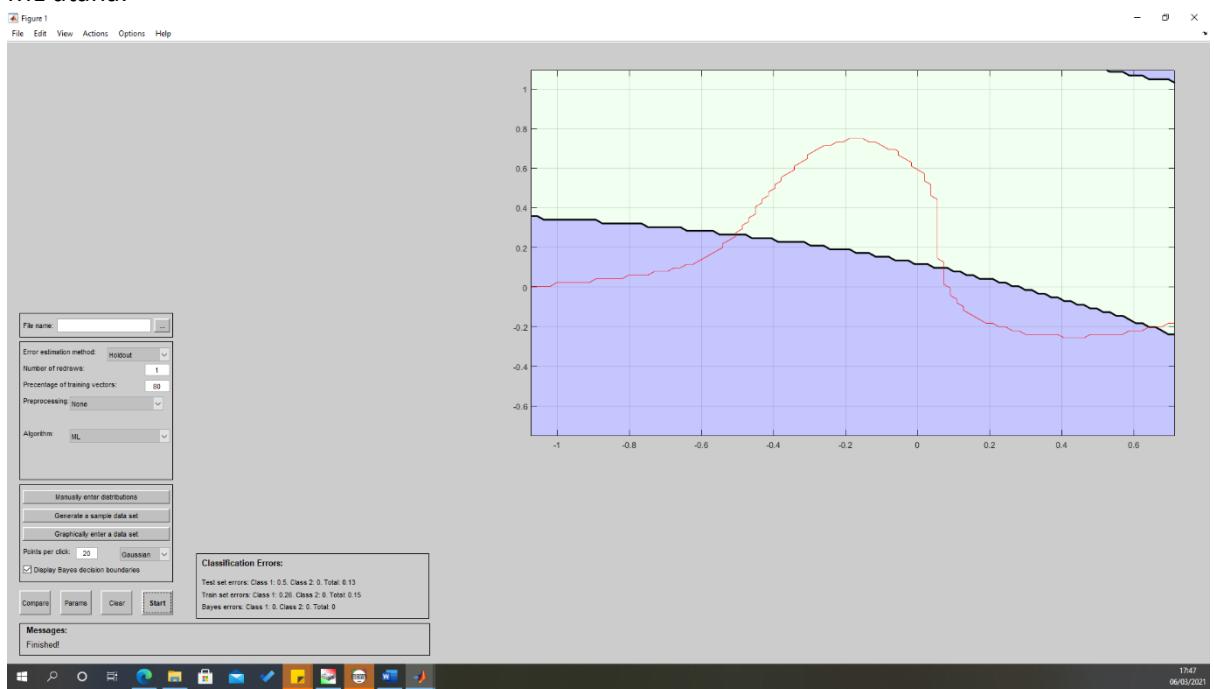
2. Saját adatbazis

A futásról készült kepernyőképek (Rajtuk a hibaszázalek)

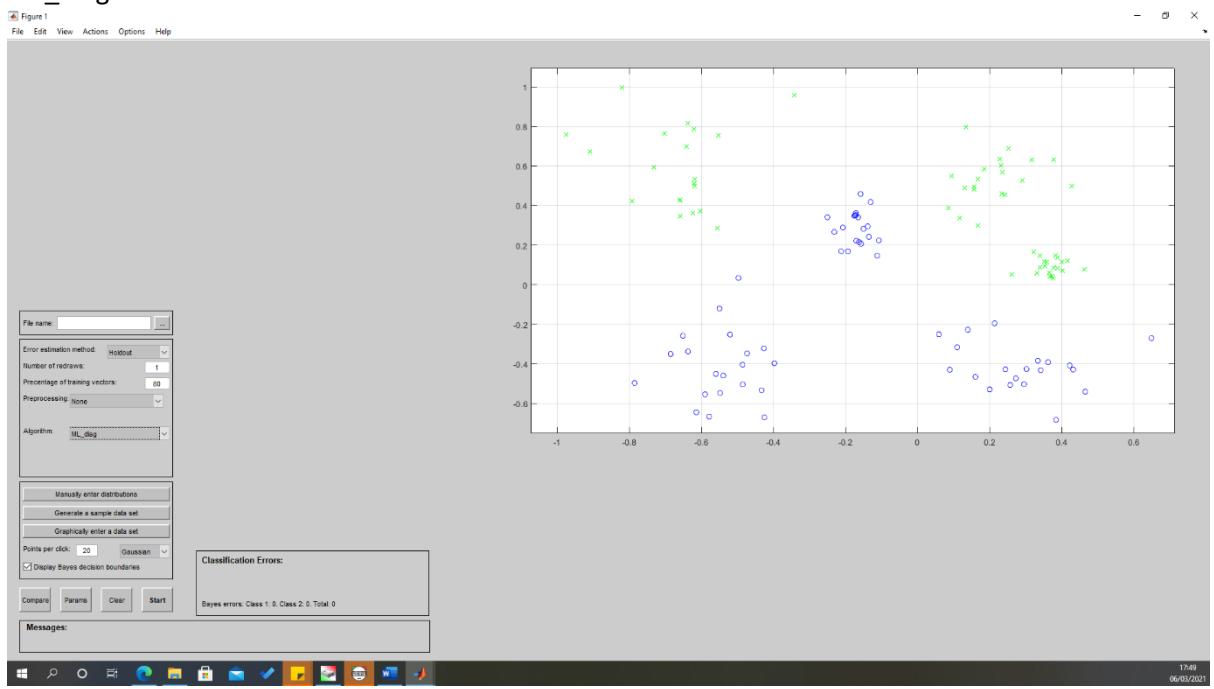
1. ML előtte:



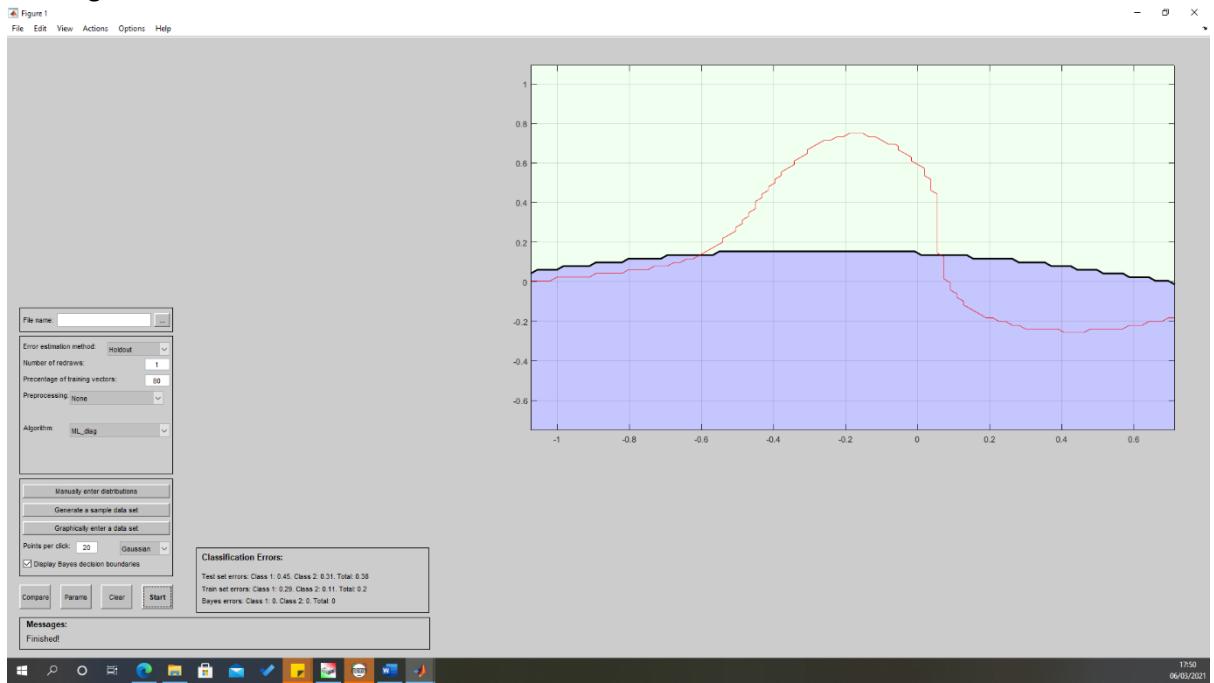
ML utana:



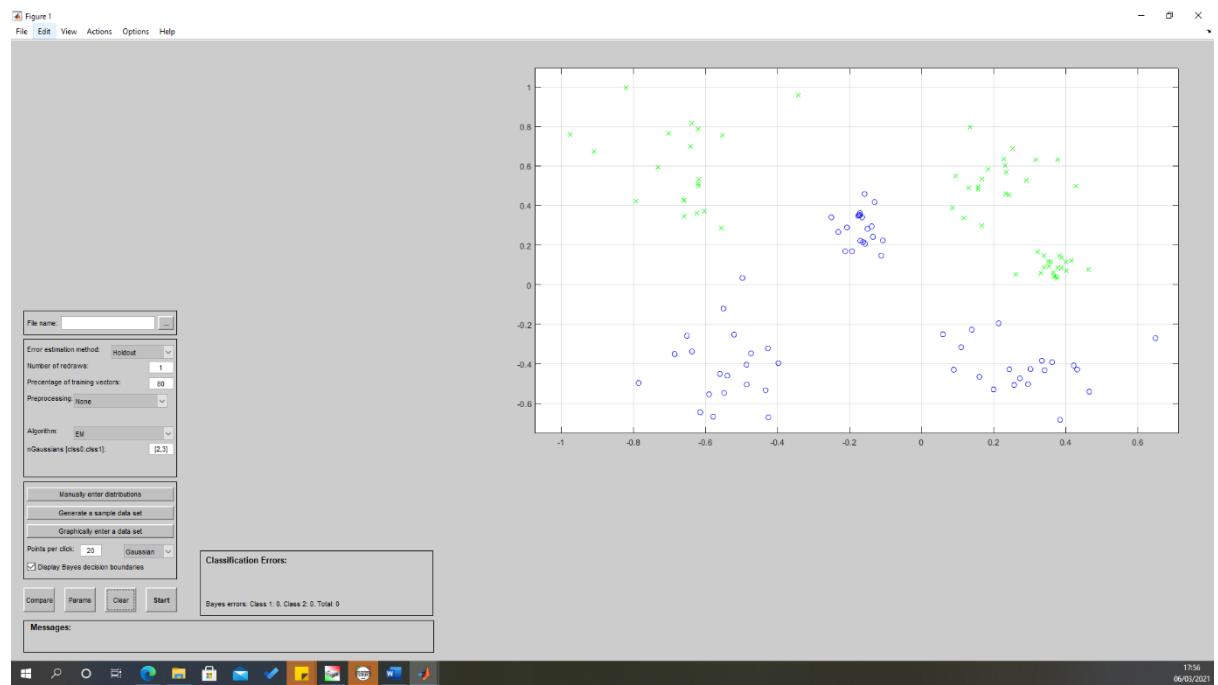
2. ML_diag előtte:



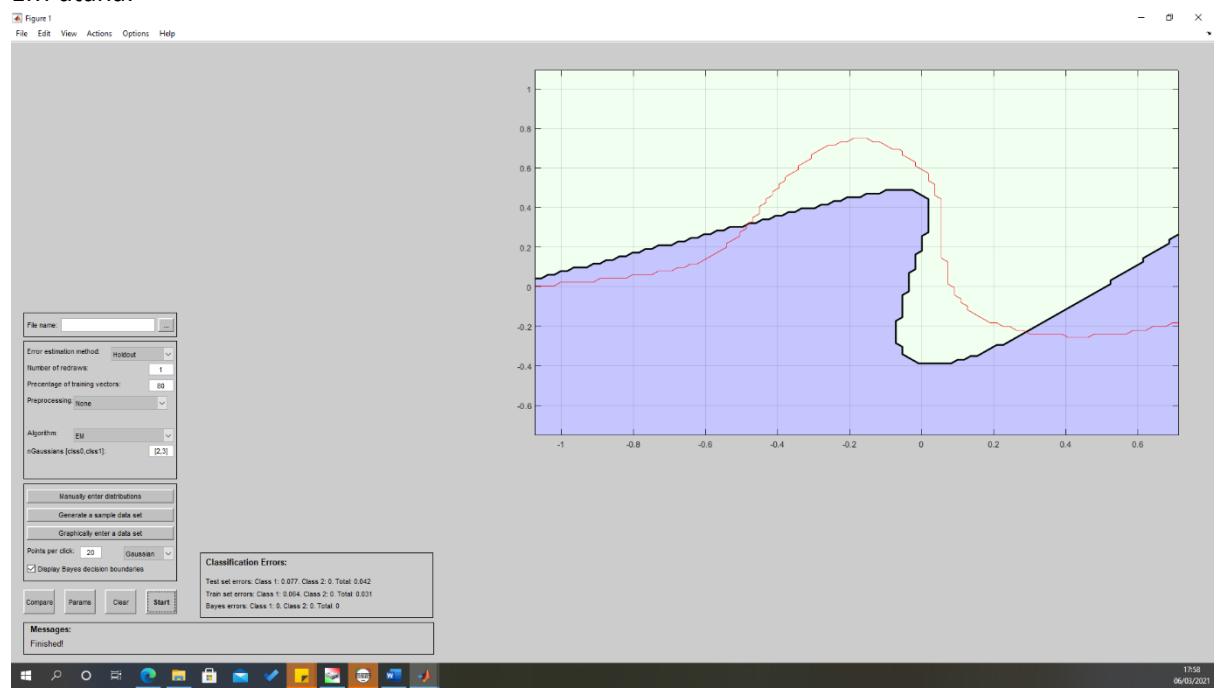
ML diag utána:



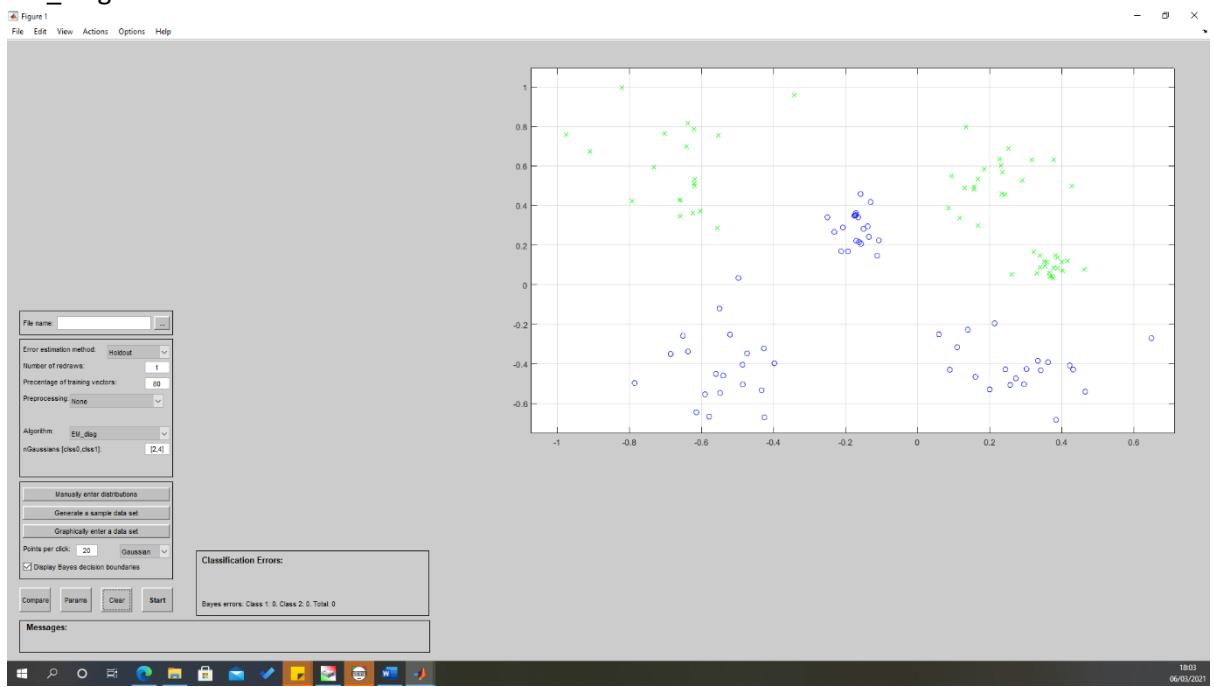
3. EM előtte:



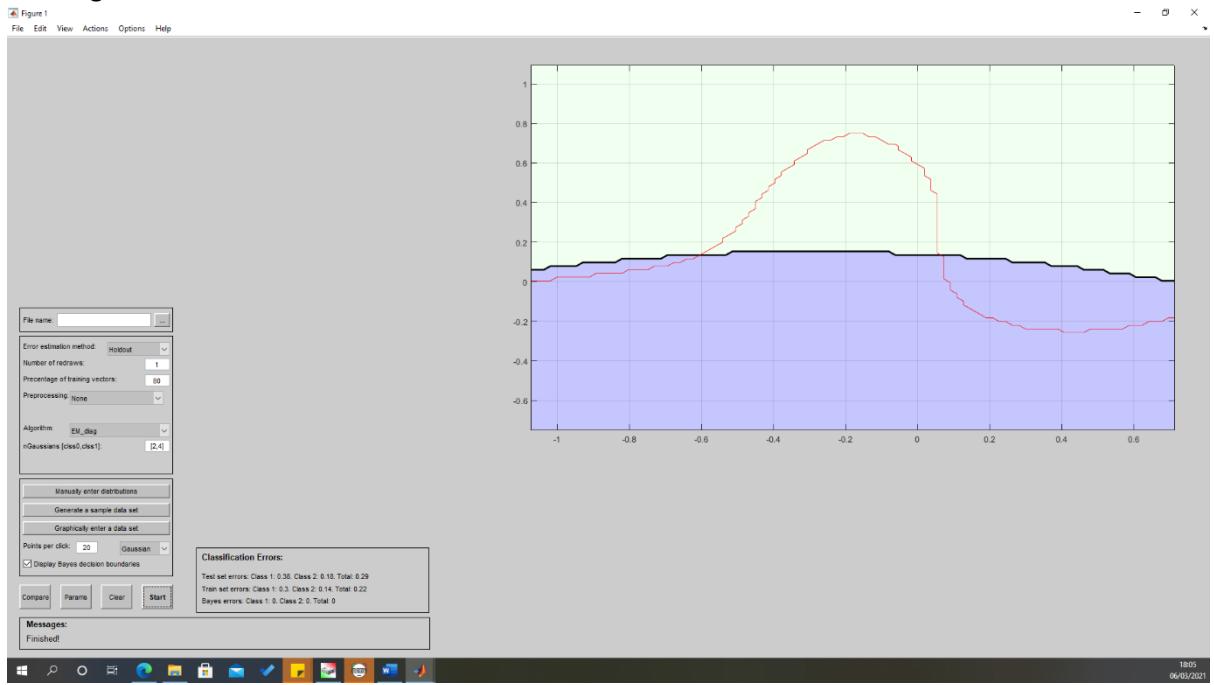
EM utana:



4. EM_diag előtte:

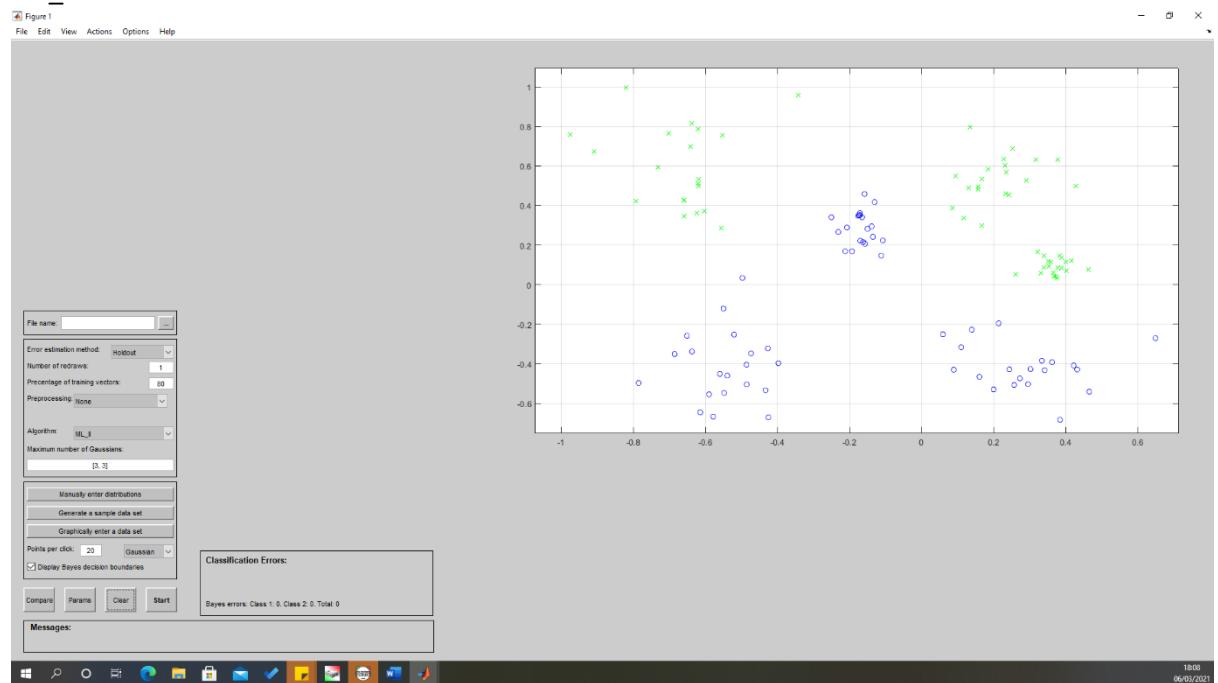


EM diag utána:

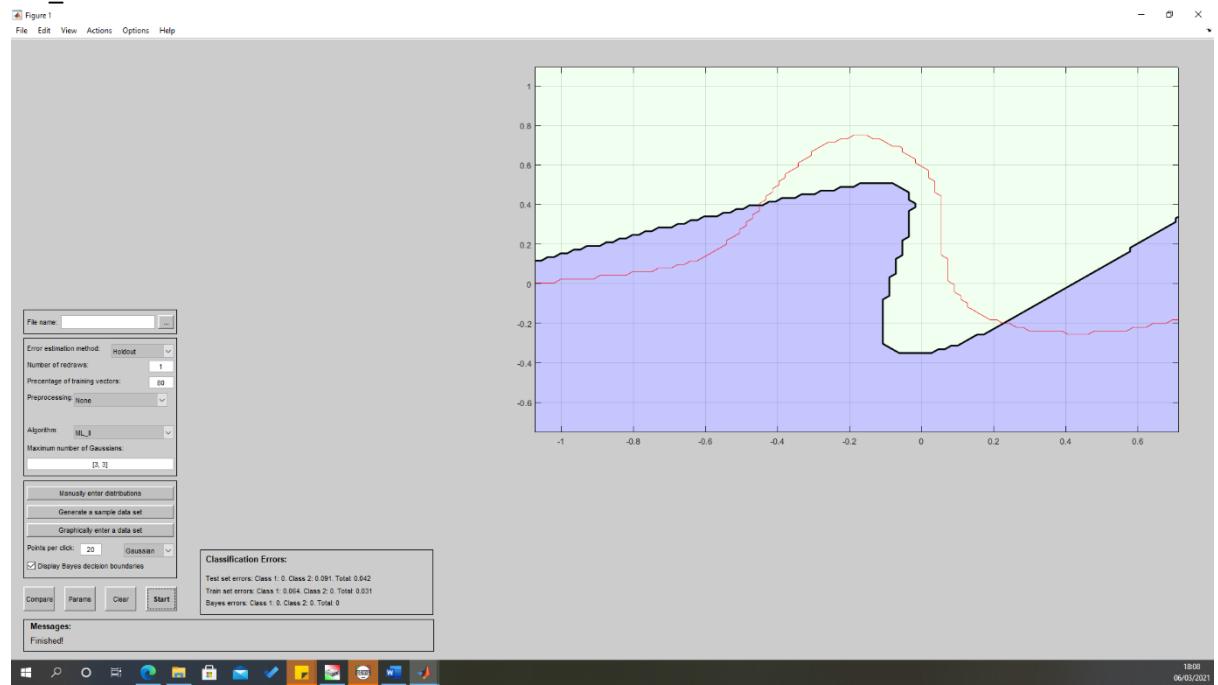


(Ennel jobb illesztést sajnos nem sikerült elernem.)

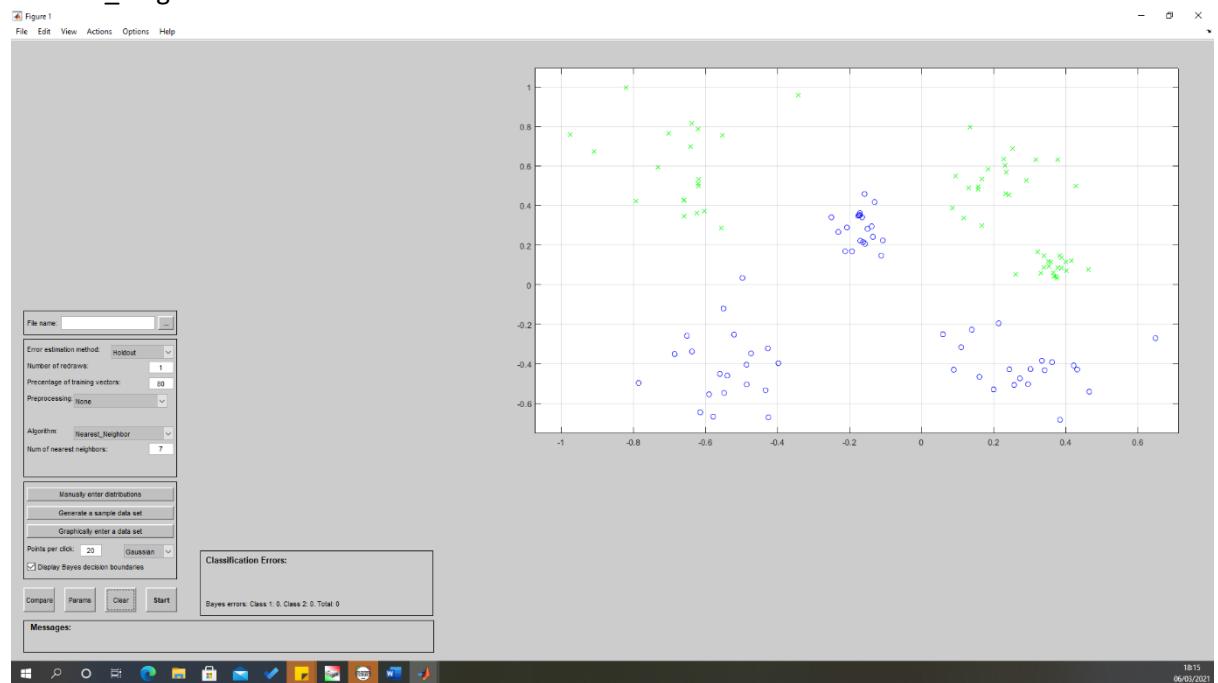
5. ML_II előtte:



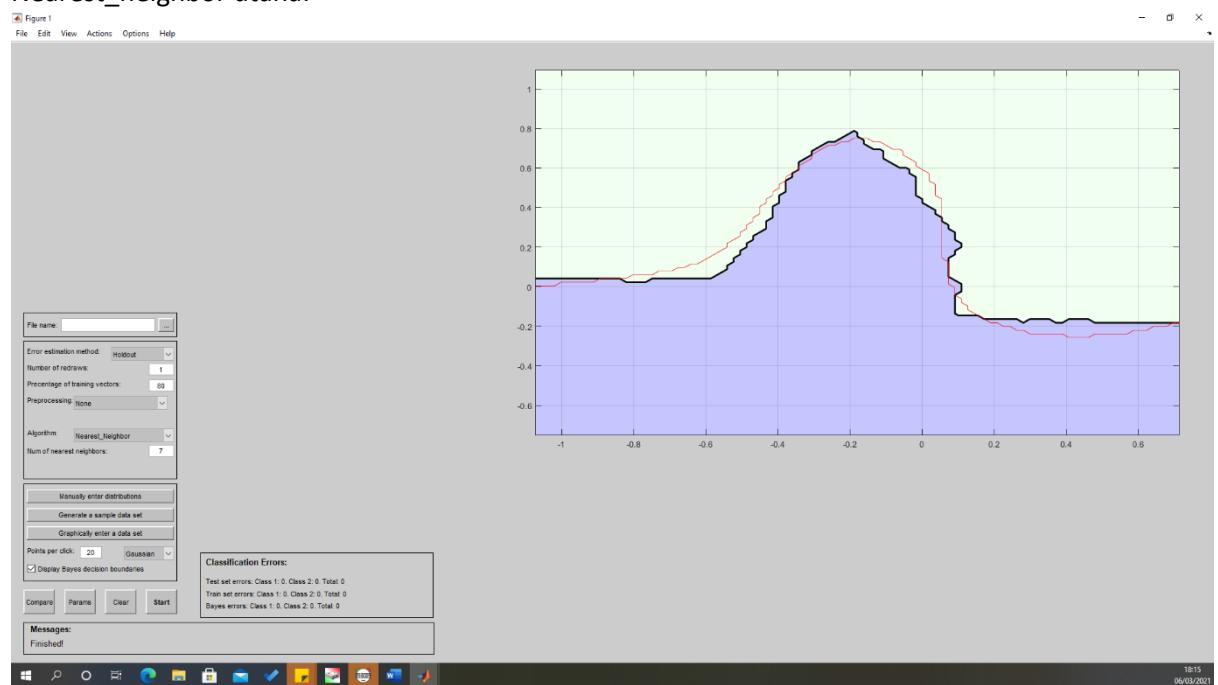
ML_II utána:



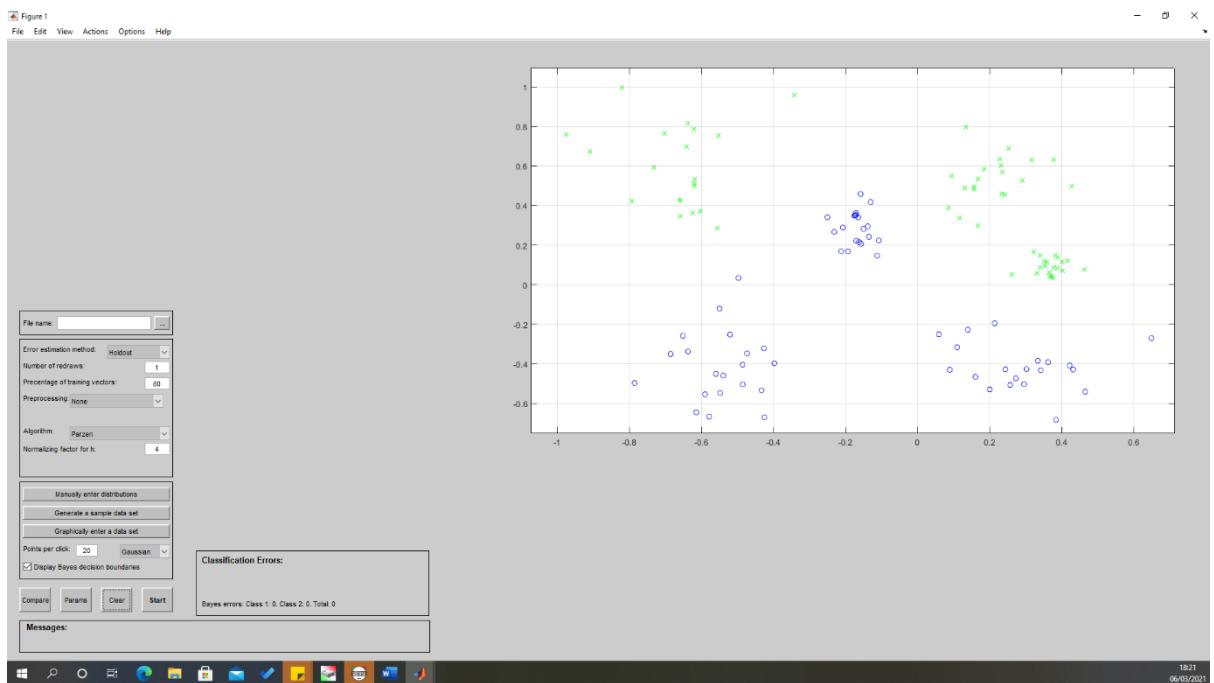
6. Nearest_neighbor elotte:



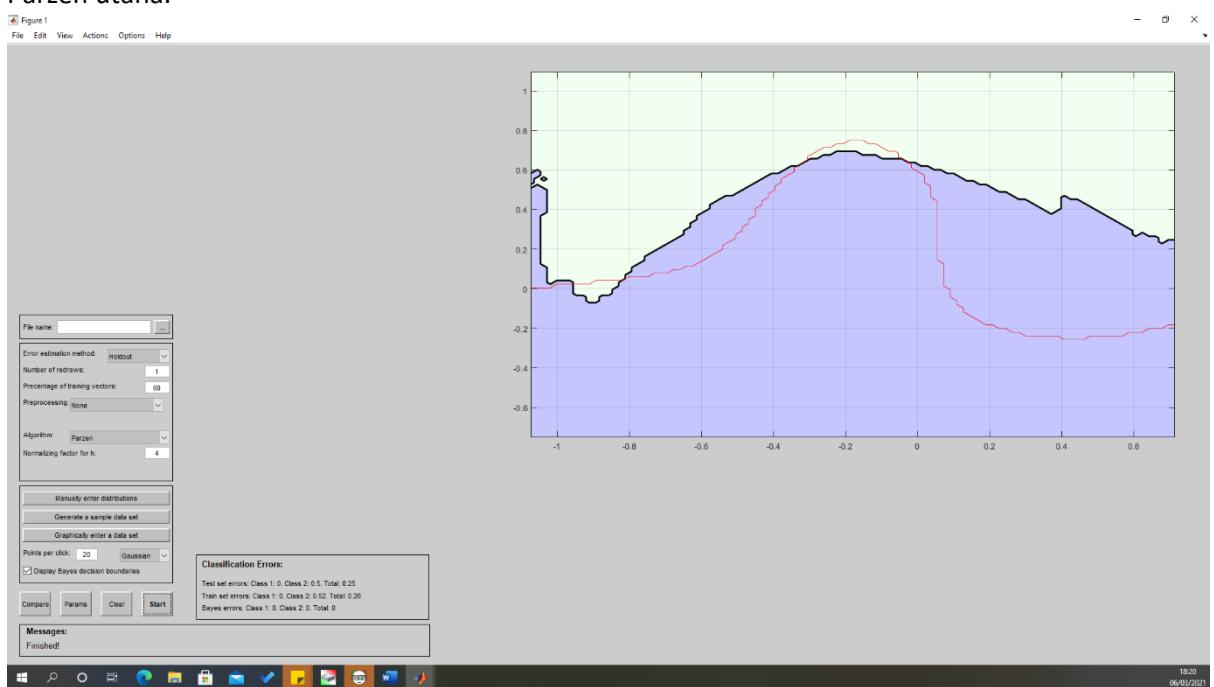
Nearest_neighbor utána:



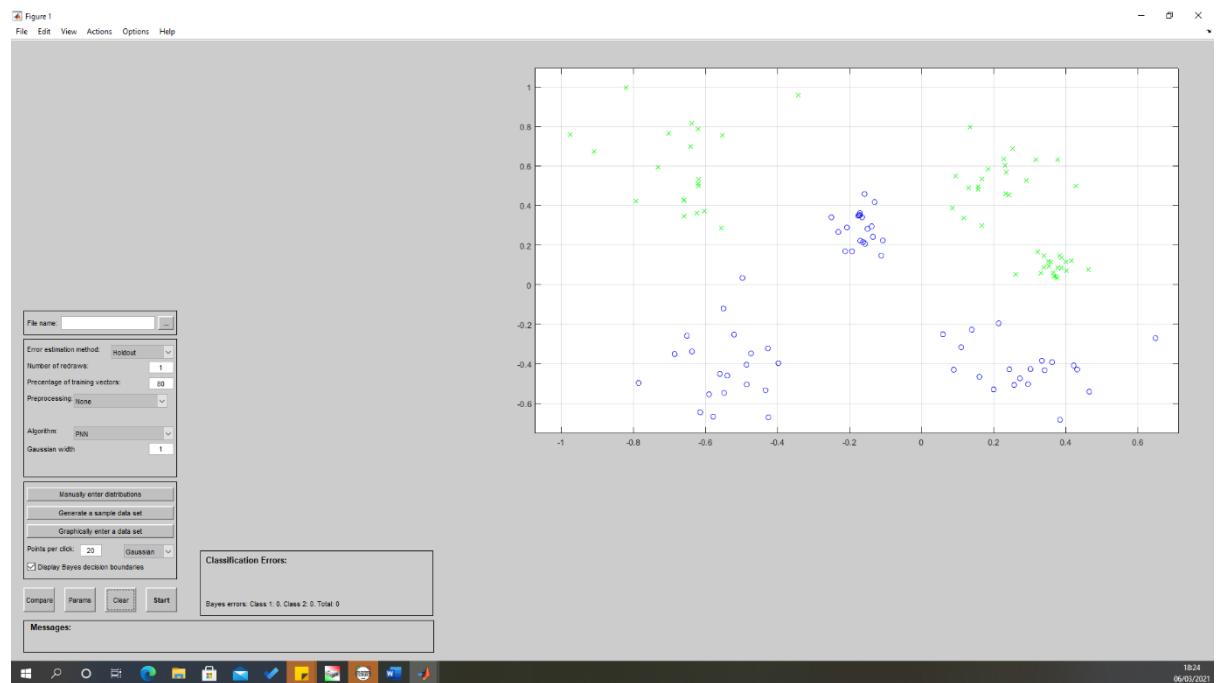
7. Parzen elotte:



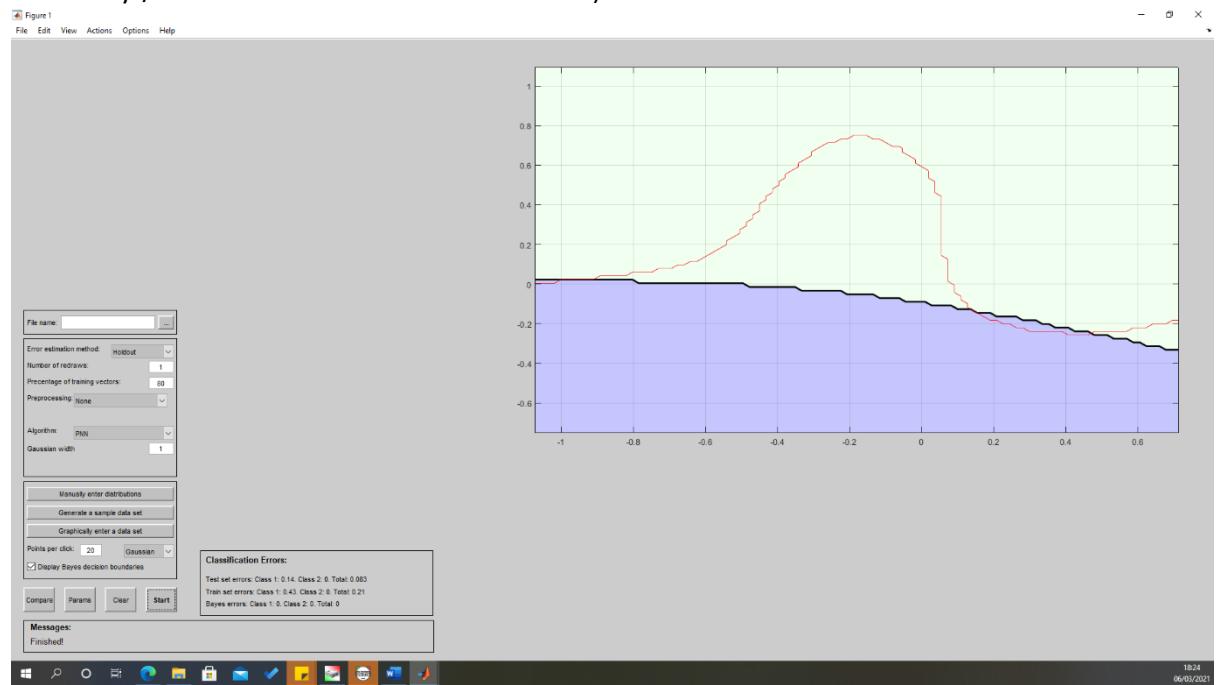
Parzen utana:



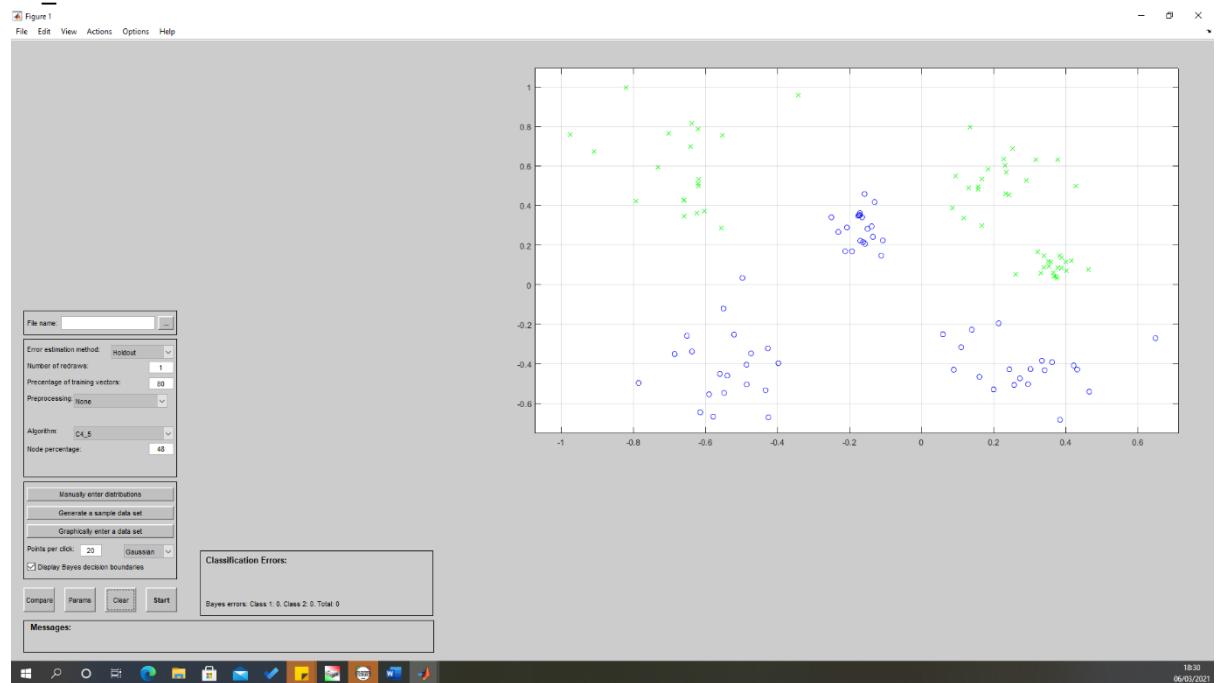
8. PNN előtte:



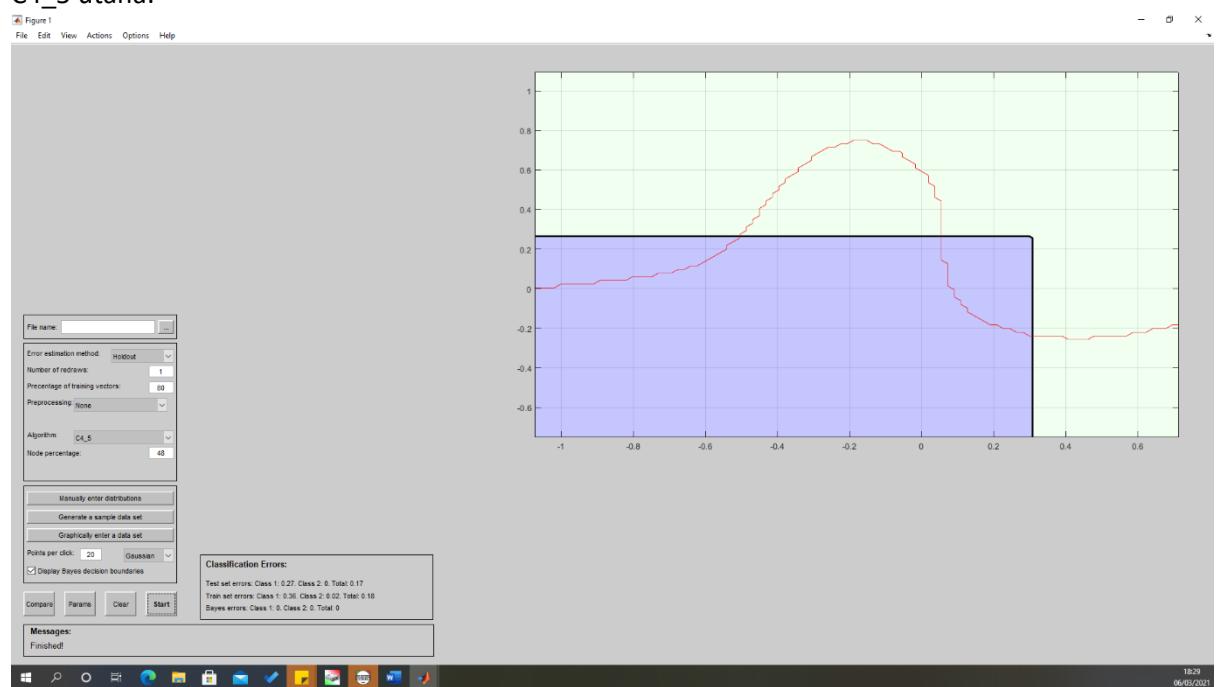
PNN utána (az alaperteknél nagyobb Gaussian width nem adott ertekelheto eredményt/sokkal rosszabb hibaszázalekot adott):



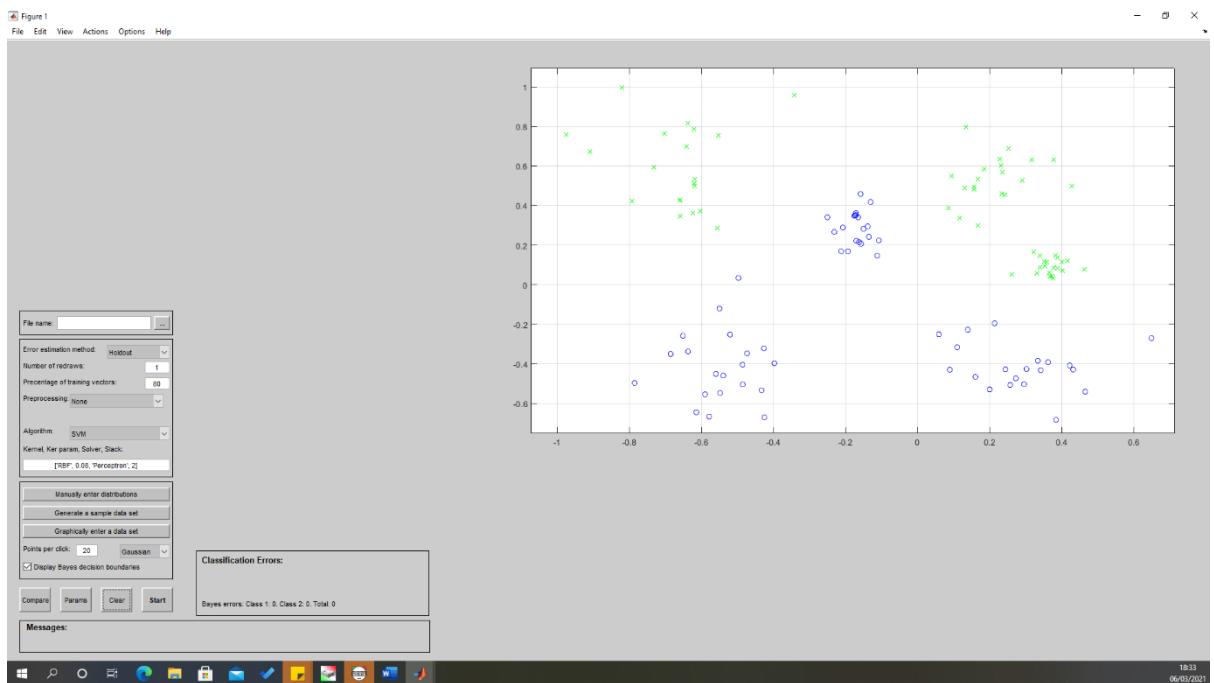
9. C4_5 előtte:



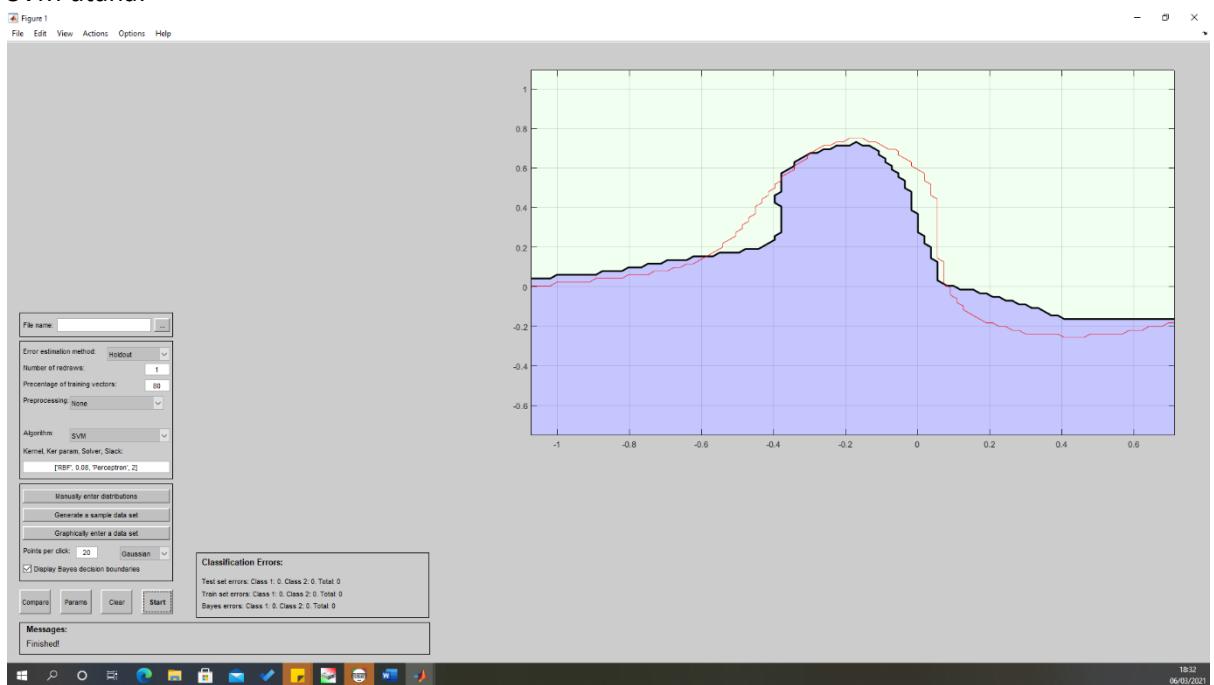
C4_5 utána:



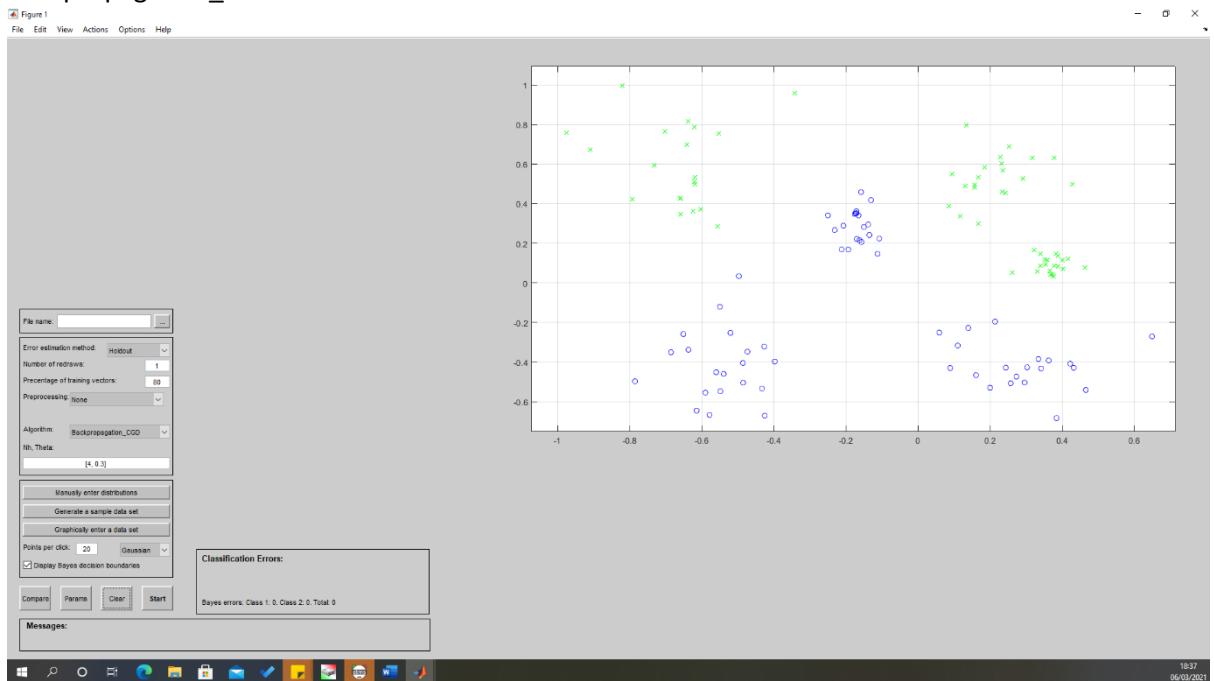
10. SVM előtte:



SVM utána:



11. Backpropagation_CGD előtte:



Backpropagation_CGD utána:

